

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS QUÍMICAS

Departamento de Bioquímica y Biología Molecular I



TESIS DOCTORAL

**Mecanismos moleculares de degradación de la pared celular vegetal:
estructura y especificidad**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

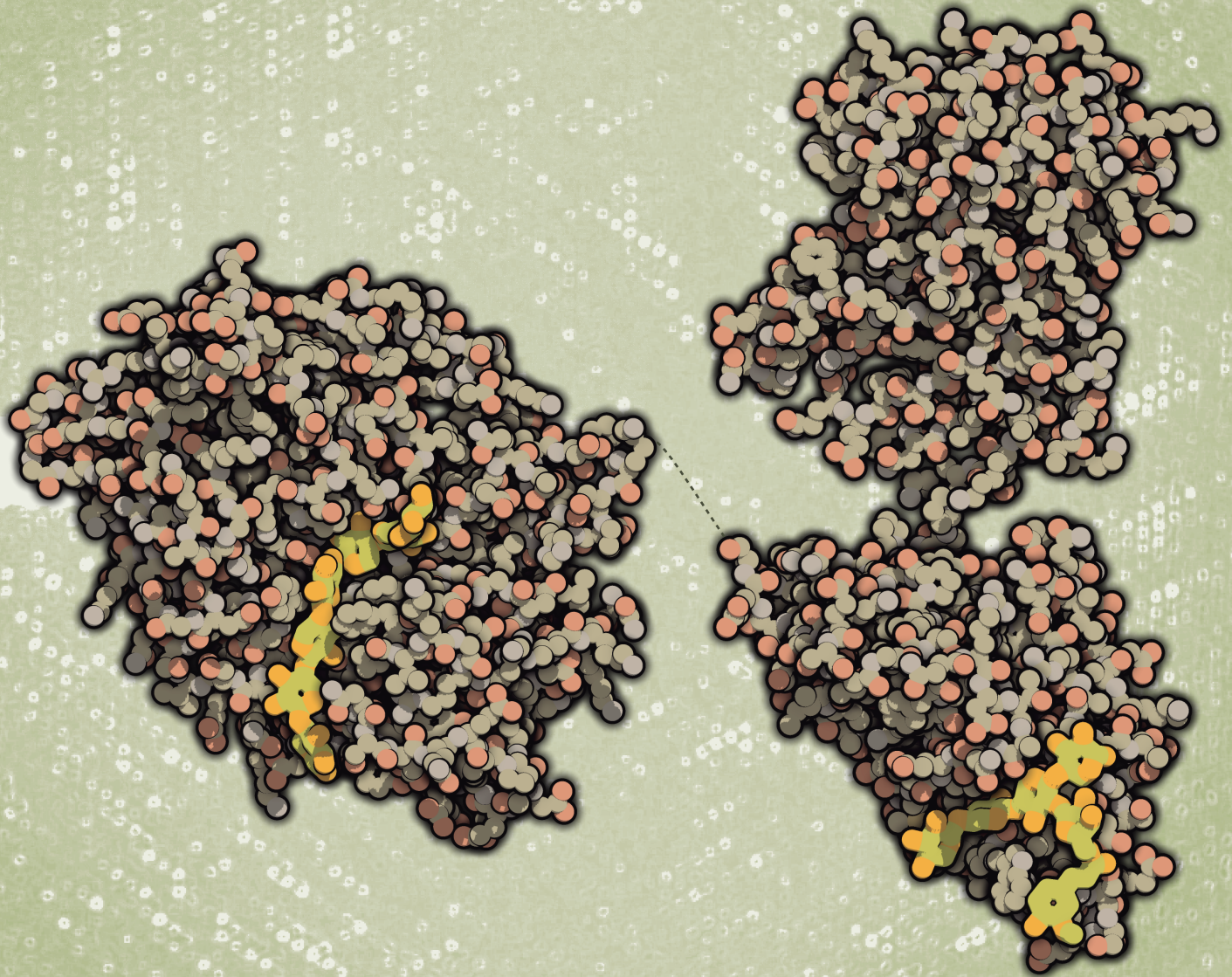
PRESENTADA POR

María Ángela Sainz Polo

Directora
Juliana Sanz Aparicio

Madrid, 2016

Mecanismos moleculares de degradación de la pared celular vegetal: estructura y especificidad de xilanasas multidominio



María Ángela Sainz Polo - Madrid, 2015

Universidad Complutense de Madrid
Consejo Superior de Investigaciones Científicas



UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS QUÍMICAS
DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR I



TESIS DOCTORAL

Mecanismos moleculares de degradación de la pared celular vegetal: estructura y especificidad de xilanasas multidominio

Memoria para optar al grado de doctor

Presentada por:
María Ángela Sainz Polo

Dirigida por:
Dra. Juliana Sanz Aparicio

DEPARTAMENTO DE CRISTALOGRAFÍA Y BIOLOGÍA ESTRUCTURAL
INSTITUTO DE QUÍMICA FÍSICA "ROCASOLANO"
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Madrid, 2015

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS QUÍMICAS

DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR I



TESIS DOCTORAL

Mecanismos moleculares de degradación de la pared celular vegetal: estructura y especificidad de xilanasas multidominio

Memoria presentada por María Ángela Sainz Polo para optar al grado de Doctor.

Esta Tesis Doctoral ha sido realizada en el Departamento de Cristalografía y Biología Estructural del Instituto de Química Física "Rocasolano" (Consejo Superior de Investigaciones Científicas, CSIC) bajo la dirección de la Dra. Juliana Sanz Aparicio, Investigador Científico del Departamento de Cristalografía y Biología Estructural del Instituto de Química Física "Rocasolano", CSIC (Consejo Superior de Investigaciones Científicas).

Madrid, 2015

La Dra. Juliana Sanz Aparicio, Investigador Científico del Departamento de Cristalografía y Biología Estructural del Instituto de Química Física “Rocasolano” (Consejo Superior de Investigaciones Científicas, CSIC), certifica que:

Doña María Ángela Sainz Polo, Licenciada en Química por la Universidad de Burgos, ha realizado bajo mi dirección el trabajo titulado “Mecanismos moleculares de degradación de la pared celular vegetal: estructura y especificidad de xilanasas multidominio”, en el Departamento de Cristalografía y Biología Estructural del Instituto de Química Física “Rocasolano” (Consejo Superior de Investigaciones Científicas, CSIC).

En mi opinión, este trabajo reúne las condiciones exigidas por la legislación vigente y tiene la originalidad, el rigor y la calidad científica necesarios y suficientes para ser presentado por Doña María Ángela Sainz Polo para optar al Grado de Doctor en Química por la Universidad Complutense de Madrid.

Para que así conste y con los efectos oportunos, firmo el presente certificado.

En Madrid, a 17 de septiembre de 2015

Fdo. Dra. Juliana Sanz Aparicio
Investigador Científico CSIC

Esta tesis fue realizada gracias a la concesión a María Ángela Sainz Polo de una beca predoctoral del CSIC JAE-Predoc en el marco del programa “Junta para la Ampliación de Estudios” (JAEPRe_2010_00107) durante el periodo comprendido entre septiembre de 2010 y agosto de 2014 por parte del Ministerio de Ciencia e Innovación.

The Spanish Ministry of Science and Innovation awarded María Ángela Sainz Polo with a “JAE-Predoc” fellowship (JAEPRe_2010_00107) from September 2010 to September 2014, to develop this Thesis.

El Ministerio de Economía y Competitividad concedió a María Ángela Sainz Polo una beca para realizar una estancia breve en la Universidad de Barcelona (España) bajo la tutela del Profesor Dr. Francisco Javier Pastor, de mayo a julio, 2012.

The Ministry of Economy and Competitiveness awarded María Ángela Sainz Polo with a grant to perform a short research stay in the University of Barcelona (Spain) under the direction of Professor Dr. Francisco Javier Pastor, from May to July, 2012.

A mis padres y a mi hermana

*A mi padre, la persona más buena, inteligente,
íntegra y valiente que he conocido*

Papá, tu fuerza es mi fuerza. Siempre a mi lado

***"Si nada nos salva de la muerte,
al menos que el amor nos salve de la vida"***

Pablo Neruda (Chile, 1904-1973)
Escritor. Premio Nobel de Literatura en 1971

Resumen

El uso de enzimas en procesos de química verde está teniendo cada vez más relevancia debido a su potencial como biocatalizadores. Así, su presencia en la industria ha crecido notablemente convirtiendo la biotecnología de proteínas en un campo emergente.

Las xilanasas son glicosil hidrolasas (GH) encargadas de la degradación del xilano, componente mayoritario de la hemicelulosa. Según su arquitectura molecular, se pueden catalogar en xilanasas de dominio simple o multidominio presentando, además del dominio catalítico (CD), uno o más módulos de unión a carbohidratos (CBM).

Normalmente, la producción de xilanasas la realizan microorganismos que durante su crecimiento están en contacto con sustratos que contienen xilano. Un ejemplo de ello es la bacteria *Paenibacillus barcinonensis*, la cual fue aislada a partir de muestras recogidas en el suelo de un arrozal en el delta del río Ebro. En este trabajo se presentan los resultados derivados del estudio estructural y funcional de dos xilanasas modulares de esta bacteria, Xyn10C y Xyn30D.

Xyn10C tiene una estructura de dominios compleja (CBM22-1-CBM22-2-GH10-CBM9a-CBM9b). En el trabajo se presenta la estructura tridimensional del dominio N-terminal Xyn10C-XBD, de cada uno de sus dos módulos aislados, de CBM22-2 con ligandos, además del dominio catalítico (Xyn10C-CD). Hay que resaltar que es la primera estructura reportada que contiene dos módulos CBM22 contiguos en tándem.

Los resultados muestran que Xyn10C-XBD se pliega en dos módulos independientes cuyo segmento conector le proporciona una amplia variedad conformacional. Además, la estructura cristalográfica de Xyn10C-CD revela que este dominio parece ser capaz de unir xilanos altamente sustituidos, pudiendo alojar ramificaciones en casi todos los subsitos.

Ambos módulos CBM22 contienen los tres residuos aromáticos del motivo de unión a sustrato (YYY) y dos residuos cargados (RE), esenciales para el reconocimiento de xilano. Los resultados estructurales nos permiten identificar un bucle presente en CBM22-2 que contiene un residuo aromático adicional, Trp308, el cual amplía la plataforma hidrofóbica y genera nuevos sitios de unión no observados anteriormente en la familia CBM22, lo que sugiere que CBM22-2 defina una subfamilia con especificidad a sitios particulares del xilano.

Según los ensayos de unión, sólo el dominio C-terminal de Xyn10C tiene la capacidad de unir celulosa, mientras que el dominio N-terminal presenta afinidad únicamente por xilano. Los geles de retardo indicaron una mayor interacción de Xyn10C-XBD frente a arabinoxilanos, presentando CBM22-2 una mayor afinidad con sustratos más sustituidos. Los ensayos de calorimetría son consistentes con los datos estructurales y de unión, mostrando dos sitios de

distinta especificidad en el tándem. Por tanto, los resultados estructurales y bioquímicos obtenidos permiten concluir que la repetición de módulos CBM de la misma familia en una proteína no se debe meramente a un efecto de multivalencia, sino que puede introducir nuevas funcionalidades enzimáticas.

La segunda proteína en estudio es Xyn30D, que presenta una composición de dominios GH30-CBM35 y es la primera xilanasa bimodular de la familia GH30 cuya estructura ha sido reportada. Además, también se describe la estructura de su CBM35 con ácidos glucurónico y aldurónico.

El estudio cristalográfico de Xyn30D muestra que ambos dominios se pliegan de forma independiente y con una flexibilidad moderada. Debido a la imposibilidad de obtener complejos de la enzima completa, ambos dominios se expresaron de forma aislada, y se obtuvieron cristales de CBM35.

Cabe destacar que únicamente se obtuvieron los complejos de CBM35 con ligandos que contenían ácido glucurónico, lo que sugiere que este es un determinante específico de unión, además esta unión es dependiente de calcio. El análisis estructural permitió identificar características únicas. Por ejemplo, el residuo Glu129 es clave en la afinidad y además destacan dos tirosinas del sitio de unión, Tyr34 y Tyr48, lo que apunta a la existencia de subsitios adicionales a los descritos anteriormente. Por último, hemos encontrado que el calcio no es imprescindible en la unión a glucuronoarabinosilano (GAX), lo que sugiere que CBM35 podría establecer enlaces con otros componentes del xilano como, por ejemplo, con arabinosa.

El estudio cristalográfico de CBM35 mostró una hendidura muy abierta con una distribución de carga negativa continua que se extiende a lo largo del sitio de unión. El modelado de un xilooligosacárido con un patrón de sustitución típico de GAX reveló una gran complementariedad con la topología de la proteína. El análisis de todos los datos obtenidos del dominio catalítico y del dominio suplementario nos ha permitido formular una hipótesis que explicaría la estrategia de actuación de Xyn30D y el papel que jugaría cada uno de sus dominios en la degradación del xilano en el hábitat natural de la bacteria.

En resumen, los resultados de este trabajo han permitido analizar los determinantes moleculares de reconocimiento proteína-carbohidrato en enzimas involucradas en la degradación de la pared celular vegetal y avanzar en el entendimiento de los sofisticados mecanismos concertados de regulación de la especificidad por los dominios adicionales no catalíticos. Se han desvelado nuevos determinantes de unión dentro de las familias estudiadas y se han establecido posibles vías para el mecanismo de actuación de dichas enzimas en el contexto general de la degradación de la pared celular vegetal.

Summary

Nowadays, green chemistry gains more and more significance and the use of enzymes within different processes also plays an increasingly important role due to their potential as biocatalysts. Not surprisingly, their presence in the industry is expanding, making enzymatic biotechnology a fast growing field.

Xylanases belong to the class of glycosyl hydrolases (GHs) which are responsible for the degradation of xylan, the major component of hemicellulose. According to their molecular architecture, they can be classified into two types: single domain or multidomain enzymes, which have one or more carbohydrate binding modules (CBMs) linked to their catalytic domain (CD).

Usually, xylanases are expressed by microorganisms that are in contact with the xylan substrate, such as *Paenibacillus barcinonensis*, a bacterium isolated from a rice field in Ebro's delta river (Spain). This particular bacterium expresses amongst others two modular xylanases, Xyn10C and Xyn30D. Here, we present the studies leading to the structural and functional elucidation of these two enzymes.

Xyn10C is a multimodular enzyme with a complex architecture (CBM22-1-CBM22-2-GH10-CBM9a-CBM9b). Our work focuses on the three dimensional structure of the N-terminal CBM22-1-CBM22-2 domain (Xyn10C-XBD), the structure of each isolated CBM22 module and the CBM22-2 complexes and on the structure of the catalytic domain (Xyn10C-CD). It should be noted that this is indeed the first reported structure containing two adjacent CBM22 modules in tandem.

The results show that Xyn10C-XBD is folded into two separated CBM22 modules, which are linked through a flexible segment providing a high plasticity to the CBM22 tandem. Furthermore, crystallographic structure elucidation of Xyn10C-CD reveals that this domain is able to bind to highly substituted xylans accommodating branches in almost all subsites.

Both CBM22 modules contain the three key aromatic residues of the substrate binding motif (YYY) and the two charged residues (RE) essential for recognition of xylan. However, structural analysis allowed us to identify a loop present in CBM22-2, which includes an additional aromatic residue, Trp308, extending the hydrophobic platform and generating new binding sites not seen previously in this family. This extra binding determinant defines a new subfamily within CBM22, directed to particular subsites of xylan.

According to the binding assays, only the C-terminal domain of Xyn10C has the ability to bind to cellulose, while the N-terminal Xyn10C domain binds specifically to xylan. These findings are underlined by results from affinity gel electrophoresis suggesting that Xyn10C-XBD binds to arabinoxylans more tightly, which is more evident when CBM22-2 is tested against highly

substituted xylan. The isothermal titration calorimetry experiments are consistent with structural and binding results showing two sites with different affinities in the tandem. Therefore, the results here presented contribute to illustrate that the presence of homologous CBM domains in tandem within the same protein sequence implies not only a multivalency effect but can also introduce new enzymatic functions.

In contrast, Xyn30D is a modular enzyme containing a family GH30 catalytic domain attached to a CBM35 module. We present here the first three-dimensional structure of a full-length bimodular protein belonging to GH30 family. In addition, the ancillary Xyn30D-CBM35 domain has been expressed and crystallized and its binding abilities have been investigated by soaking experiments.

Crystallographic studies of Xyn30D show that, although both domains fold in an independent manner, the linker region makes polar interactions with the catalytic domain, allowing a moderate flexibility. Due to the inability to obtain complexes from the full-length enzyme, both domains were expressed independently but only CBM35 crystallized. The CBM35 crystals were then used to obtain complexes with several ligands. Interestingly, only CBM35 complexes with glucuronic acid-containing ligands were obtained, suggesting that this is a specific binding determinant. Moreover, the structural analysis allowed the identification of the unique structural features regulating substrate binding: Glu129 plays a key role in binding while two residues, Tyr34 and Try48, are novel elements that extend the binding site. Surprisingly, calcium is not essential for glucuronoarabinoxylan (GAX) binding, which could suggest that CBM35 might interact with other xylan components, such as arabinose.

The main feature observed in the Xyn30D-CBM35 structure is a very open crevice with a negative charge distribution that extends along its binding pocket. A xylooligosaccharide chain with a decoration pattern typical of GAX found in grass cell walls has been modeled revealing a great complementarity with the protein topology. The analysis of all data allowed us to formulate a hypothesis that would explain the role of Xyn30D in the degradation of xylan in the natural habitat of the bacterium.

In summary, the results presented here extend our knowledge on the molecular determinants of protein-carbohydrate recognition of enzymes involved in degradation of plant cell wall. Also, they contribute to improve the understanding of the sophisticated mechanisms displayed by the ancillary non-catalytic domains in regulating specificity. Novel binding determinants have been uncovered within the enzymes studied, which pave the way to define new strategies inside the general context of the degradation of plant cell wall.

Listado de publicaciones

El trabajo de investigación presentado en esta Tesis está recogido en las siguientes publicaciones:

1. **Crystallization and preliminary X-ray diffraction analysis of the N-terminal domain of *Paenibacillus barcinonensis* xylanase 10C containing the CBM22-1-CBM22-2 tandem.**
Sainz-Polo, M. A., González, B., Pastor, F. I. & Sanz-Aparicio, J. (2015). *Acta Crystallographica. Section F, Structural Biology Communications* **71**, 136-140.
2. **Exploring multimodularity in plant cell wall deconstruction: Structural and functional analysis of Xyn10C containing the CBM22-1-CBM22-2 tandem.**
Sainz-Polo, M. A., González, B., Menéndez, M., Pastor, F. I. & Sanz-Aparicio, J. (2015). *The Journal of Biological Chemistry* **290**, 17116-17130.
3. **Crystallization and preliminary X-ray diffraction analysis of Xyn30D from *Paenibacillus barcinonensis*.**
Sainz-Polo, M. A., Valenzuela, S. V., Pastor, F. J. & Sanz-Aparicio, J. (2014). *Acta Crystallographica. Section F, Structural Biology Communications* **70**, 963-966.
4. **Structural analysis of glucuronoxylan-specific Xyn30D and its attached CBM35 domain gives insights into the role of modularity in specificity.**
Sainz-Polo, M. A., Valenzuela, S. V., González, B., Pastor, F. I. & Sanz-Aparicio, J. (2014). *The Journal of Biological Chemistry* **289**, 31088-31101.

Paralelamente, se han realizado estudios con otras glicosil hidrolasas que intervienen en procesos biológicos distintos pero que presentan mecanismos catalíticos similares. Dichos trabajos han generado las siguientes publicaciones:

5. **Three-dimensional structure of *Saccharomyces* invertase: Role of a non-catalytic domain in oligomerization and substrate specificity.**
Sainz-Polo, M. A., Ramírez-Escudero, M., Lafraya, A., González, B., Marín-Navarro, J., Polaina, J. & Sanz-Aparicio, J. (2013). *The Journal of Biological Chemistry* **288**, 9755-9766.
6. **Crystallization and preliminary X-ray diffraction analysis of the invertase from *Saccharomyces cerevisiae*.**
Sainz-Polo, M. A., Lafraya, A., Polo, A., Marín-Navarro, J., Polaina, J. & Sanz-Aparicio, J. (2012). *Crystallographica. Section F, Structural Biology Communications* **68**, 1538-1541.

7. **Structural and kinetic insights reveal that the amino acid pair Gln-228/Asn-254 modulates the transfructosylating specificity of *Schwanniomyces occidentalis* β -fructofuranosidase, an enzyme that produces prebiotics.**

Álvaro-Benito, M., Sainz-Polo, M. A., González-Pérez, D., González, B., Plou, F. J., Fernández-Lobato, M. & Sanz-Aparicio, J. (2012). *The Journal of Biological Chemistry* **287**, 19674-19686.

Abreviaturas

| | |
|---------------|--|
| adj. | Adjetivo |
| ADN | Ácido desoxirribonucleico (<i>Deoxyribonucleic acid</i>) |
| AGE | Electroforesis en gel de afinidad (<i>Affinity gel electrophoresis</i>) |
| AX | Arabinoxilano |
| AXOS | Xilooligosacáridos procedentes del arabinoxilano |
| Biol. | Biología |
| Bioquím. | Bioquímica |
| Bis-tris | Bis(2-hidroxietil)-amino-tris(hidroximetil)-metano |
| Bot. | Botánica |
| BSA | Seroalbúmina bovina (<i>Bovine serum albumin</i>) |
| CAZy | Enzimas activas en carbohidratos (<i>Carbohydrate-active enzymes</i>) |
| CBD | Dominio de unión a celulosa (<i>Cellulose binding domain</i>) |
| CBM | Módulo de unión a carbohidrato (<i>Carbohydrate binding module</i>) |
| CC | Coeficiente de correlación de Pearson |
| CCD | Dispositivo de carga acoplada (<i>Charged-coupled device</i>) |
| CCP4 | Collaborative Computacional Project, number 4 |
| CD | Módulo catalítico (<i>Catalytic domain</i>) |
| cDNA | ADN complementario (<i>Complementary DNA</i>) |
| CE | Carbohidrato esterasa |
| Da | Dalton |
| DESY | Sincrotrón alemán de electrones (<i>Deutsches elektronen synchrotron</i>) |
| DO | Densidad óptica |
| DP | Grado de polimerización (<i>Degree of polymerization</i>) |
| EC | Comisión de enzimas (<i>Enzyme commission</i>) |
| EDTA | Ácido etildiaminotetraacético (<i>Ethylenediaminetetraacetic acid</i>) |
| ESRF | Instalación europea de radiación sincrotrón (<i>European synchrotron radiation facility</i>) |
| <i>et al.</i> | y otros (de latín <i>et alii</i>) |
| f. | Femenino |
| FPLC | Cromatografía rápida de proteínas en fase líquida (<i>Fast protein liquid chromatography</i>) |
| g | Gramo |
| G | Unidad de centrifugación |
| GAX | Glucuronoarabinoxilano |
| GGM | Galactoglucomanano |
| GH | Glicosil hidrolasa |
| GlaA | Ácido galacturónico (<i>Galacturonic acid</i>) |
| Glc | Glucosa |
| GlcA | Ácido glucurónico (<i>Glucuronic acid</i>) |
| GM | Galactomanano |
| GS | Glucanos |
| GST | Glutación-S-transferasa |
| GX | Glucuronoxilano |
| h | Hora |
| HEPES | Ácido 4-(2-hidroxietil)-1-piperazinaetanosulfónico (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) |
| IEX | Cromatografía de intercambio iónico (<i>Ionic exchange chromatography</i>) |
| IMAC | Cromatografía de afinidad por ión metálico inmovilizado (<i>Immobilized metal ion affinity chromatography</i>) |
| IP | Placa de imagen (<i>Image plate</i>) |
| IPTG | Isopropil-β-D-tiogalactósido |
| ITC | Calorimetría isotérmica de titulación (<i>Isothermal titration calorimetry</i>) |
| K | Constante de unión |
| K | Grado Kelvin |
| kDa | Kilodalton |

| | |
|----------|---|
| I | <u>L</u> itro |
| LB | Caldo de lisogenia (<u>L</u> ysogeny <u>b</u> roth) |
| M | <u>M</u> olar |
| m. | <u>M</u> asculino |
| m/v | Relación <u>m</u> asa/ <u>v</u> olumen |
| mA | <u>M</u> iliamperio |
| MAD | Dispersión anómala múltiple (<u>M</u> ulti- <u>w</u> avelength <u>a</u> nomalous <u>d</u> ispersion) |
| MeGAX | <u>M</u> etilglucurono <u>a</u> rabino <u>x</u> ilano |
| MeGlcA | Ácido metilglucurónico (<u>M</u> ethyl <u>g</u> luc <u>u</u> ronic <u>a</u> cid) |
| mg | <u>M</u> iligramo |
| min | Minuto |
| MIR | Reemplazamiento isomorfo múltiple (<u>M</u> ultiple <u>i</u> somorphous <u>r</u> eplacement) |
| MIRAS | Reemplazamiento isomorfo múltiple con dispersión anómala (<u>M</u> ultiple <u>i</u> somorphous <u>r</u> eplacement with <u>a</u> nomalous <u>s</u> cattering) |
| mm | <u>M</u> ilímetro |
| mM | <u>M</u> ilimolar |
| MPD | 2- <u>M</u> etil-2,4-pentanediol |
| MR | Reemplazamiento Molecular (<u>M</u> olecular <u>r</u> eplacement) |
| n. | <u>N</u> ombre |
| o | <u>O</u> rto |
| °C | Grado Celsius |
| Org. | <u>O</u> rgánica |
| p | <u>P</u> ara |
| PAD | Detectores de matriz de píxeles (<u>P</u> ixel <u>a</u> rray <u>d</u> etectors) |
| PAGE | Electroforesis en gel de poliacrilamida (<u>P</u> olyacrylamide gel electrophoresis) |
| PCR | Reacción en cadena de la polimerasa (<u>P</u> olymerase <u>c</u> hain <u>r</u> eaction) |
| PCT | Test de pre-cristalización (<u>P</u> re- <u>c</u> rySTALLization <u>t</u> est) |
| PDB | Banco de datos de proteínas (<u>P</u> rotein <u>d</u> ata <u>b</u> ank) |
| PEG | <u>P</u> olietilenglicol |
| PEG 3350 | <u>P</u> olietilenglicol 3350 Da |
| PEG 4K | <u>P</u> olietilenglicol 4 kDa |
| pI | <u>P</u> unto isoeléctrico |
| PL | <u>P</u> ectato liasa |
| pl. | <u>P</u> lural |
| Quím. | <u>Q</u> uímica |
| RMN | <u>R</u> esonancia <u>m</u> agnética <u>n</u> uclear |
| rmsd | Desviación cuadrática media (<u>R</u> oot <u>m</u> ean <u>s</u> quare <u>d</u> eviation) |
| rpm | <u>R</u> evoluciones por <u>m</u> inuto |
| s | <u>S</u> egundo |
| SAD | Dispersión anómala simple (<u>S</u> ingle wavelength <u>a</u> nomalous <u>d</u> ispersion) |
| SAXS | Dispersión de rayos X a bajo ángulo (<u>S</u> mall <u>a</u> ngle <u>X</u> -ray <u>s</u> cattering). |
| SCN | <u>S</u> ulfocianuro |
| SDS | Dodecilsulfato sódico(<u>S</u> odium <u>d</u> odecyl <u>s</u> ulfate) |
| SEM | Microscopía electrónica de barrido (<u>S</u> canning electron <u>m</u> icroscope) |
| Sin. | <u>S</u> inónimo |
| SIR | Reemplazo isomorfo simple (<u>S</u> ingle <u>i</u> somorphous <u>r</u> eplacement) |
| SIRAS | Reemplazo isomorfo simple con dispersión anómala (<u>S</u> ingle <u>i</u> somorphous <u>r</u> eplacement with <u>a</u> nomalous <u>s</u> cattering) |
| SLH | Homología de capa S (<u>S</u> -layer <u>h</u> omology) |
| sp. | Especie |
| TEM | Microscopía electrónica de transmisión (<u>T</u> ransmission electron <u>m</u> icroscopy) |
| TLC | Cromatografía en capa fina (<u>T</u> hin layer chromatography) |
| Tm | Temperatura de desnaturalización (<u>M</u> elting <u>t</u> emperature) |
| Tris | <u>T</u> ris(hidroximetil)aminometano |
| U | <u>U</u> nidad de actividad |
| UA | <u>U</u> nidad <u>a</u> simétrica |

| | |
|--------------------|--|
| UV | <u>U</u> ltravioleta |
| V | <u>V</u> oltios |
| v/v | Relación <u>v</u> olumen/ <u>v</u> olumen |
| XBD | Dominio de unión a xilano (<u>X</u> ylan <u>b</u> inding <u>d</u> omain) |
| XFEL | Láser de rayos X de electrones libres (<u>X</u> -ray free <u>e</u> lectron <u>l</u> aser) |
| XG | <u>X</u> iloglucano |
| XOS | <u>X</u> ilooligosacáridos |
| Xyl | Xilano (<u>X</u> ylan) |
| Xyn | Xilanasa (<u>X</u> ylanase) |
| $\Delta 4,5$ -GalA | Ácido $\Delta 4,5$ -anhidrogalaacturónico (<u>$\Delta 4,5$</u> -anhydrogalaturonic <u>a</u> cid) |
| ΔG | Energía libre de Gibbs |
| ΔH | Entalpía de unión |
| ΔS | Entropía |
| μ l | <u>M</u> icrolitro |
| μ m | <u>M</u> icrómetro/ |
| μ M | <u>M</u> icromolar |
| μ mol | <u>M</u> icromol |
| $\times g$ | Aceleración de la gravedad |

ÍNDICE DE CONTENIDOS

| | |
|---|------------|
| Resumen | I |
| Summary | III |
| Listado de publicaciones | V |
| Abreviaturas | VII |
| Índice de contenidos | 1 |
| 1. Introducción | 5 |
| 1.1. Pared celular vegetal e importancia fisiológica | 7 |
| 1.1.1. Composición química y estructura | 9 |
| 1.1.2. Celulosa | 11 |
| 1.1.3. Hemicelulosa | 13 |
| 1.1.4. Pectina | 15 |
| 1.1.5. Lignina | 16 |
| 1.2. El xilano | 17 |
| 1.2.1. Degradación enzimática del xilano | 19 |
| 1.3. Xilanasas y su mecanismo catalítico | 20 |
| 1.3.1. Arquitectura molecular y clasificación | 24 |
| 1.3.2. Xilanasas GH10 | 27 |
| 1.3.3. Xilanasas GH30 | 28 |
| 1.4. Módulos de unión a carbohidratos | 30 |
| 1.4.1. Clasificación | 31 |
| 1.4.2. Función | 35 |
| 1.4.3. Papel del hidrógeno y del calcio | 37 |
| 1.4.4. Familias CBM22 y CBM9 | 37 |
| 1.4.5. Familia CBM35 | 39 |
| 1.4.6. Ensayos de unión <i>in vivo</i> mediante inmunohistoquímica | 41 |
| 1.5. Aplicaciones biotecnológicas de las xilanasas | 42 |
| 1.5.1. Biomedicina e industria farmacéutica | 42 |
| 1.5.2. Conversión de la biomasa | 43 |
| 1.5.3. Otros usos industriales | 43 |
| 1.6. Nuevas aplicaciones de los módulos de unión a carbohidratos | 44 |
| 1.7. <i>Paenibacillus barcinonensis</i> | 45 |
| 2. Objetivos | 47 |
| 3. Metodología Experimental | 51 |
| 3.1. Metodología en cristalografía macromolecular | 53 |
| 3.1.1. Aspectos generales | 53 |
| 3.1.2. Cristalización de proteínas | 54 |
| 3.1.2.1. Concepto de cristal | 55 |
| 3.1.2.2. Principios básicos de cristalización | 56 |
| 3.1.2.3. Requisitos previos a la cristalización | 57 |
| 3.1.2.4. Experimentos de cristalización | 58 |
| 3.1.2.4.1. Metodologías de cristalización | 58 |
| 3.1.2.4.2. Búsqueda de condiciones de cristalización | 59 |
| 3.1.2.4.3. Optimización de los cristales | 60 |
| 3.1.2.4.4. Cristalización de complejos proteína-ligando | 61 |
| 3.1.2.5. Criocristalografía | 61 |
| 3.1.3. Difracción de rayos X | 62 |
| 3.1.3.1. Introducción al fenómeno de difracción | 62 |
| 3.1.3.1.1. Fuentes de radiación de rayos X | 63 |

| | |
|---|------------|
| 3.1.3.1.2. Crioprotección | 65 |
| 3.1.3.1.3. Detector | 66 |
| 3.1.3.2. Recogida de datos de difracción..... | 67 |
| 3.1.4. Resolución estructural..... | 67 |
| 3.1.4.1. Procesado de datos | 67 |
| 3.1.4.2. Resolución del problema de la fase..... | 69 |
| 3.1.4.2.1. Método de Reemplazo Molecular | 70 |
| 3.1.4.3. Mejora de las fases | 71 |
| 3.1.4.4. Construcción del modelo estructural y refinado | 71 |
| 3.1.4.5. Validación del modelo | 73 |
| 3.2. Otras técnicas físico-químicas | 74 |
| 3.2.1. Calorimetría Isotérmica de Titulación | 74 |
| 3.2.2. Ultracentrifugación analítica | 75 |
| 3.3. Otras herramientas bioinformáticas | 75 |
| 4. Trabajos de Investigación | 77 |
| 4.1. Artículo 1:..... | 79 |
| Crystallization and preliminary X-ray diffraction analysis of the N-terminal domain of <i>Paenibacillus barcinonensis</i> xylanase 10C containing the CBM22-1-CBM22-2 tandem | 79 |
| 4.1.1. Introducción a la cristalización del dominio N-terminal, CBM22-1-CBM22-2, de la xilanasa 10C de <i>Paenibacillus barcinonensis</i> | 81 |
| 4.1.2. Publicación..... | 83 |
| 4.1.3. Resultados y discusión..... | 88 |
| 4.2. Artículo 2:..... | 89 |
| Exploring multimodularity in plant cell wall deconstruction: Structural and functional analysis of Xyn10C containing the CBM22-1-CBM22-2 tandem..... | 89 |
| 4.2.1. Introducción al análisis estructural y funcional del dominio N-terminal de la xilanasa Xyn10C de <i>Paenibacillus barcinonensis</i> | 91 |
| 4.2.2. Publicación..... | 93 |
| 4.2.3. Resultados | 109 |
| 4.2.4. Discusión | 110 |
| 4.2.5. Conclusiones: Artículos 1 y 2 | 112 |
| 4.3. Artículo 3:..... | 115 |
| Crystallization and preliminary X-ray diffraction analysis of Xyn30D from <i>Paenibacillus barcinonensis</i>..... | 115 |
| 4.3.1. Introducción a la cristalización de Xyn30D de <i>Paenibacillus barcinonensis</i> .. | 117 |
| 4.3.2. Publicación..... | 119 |
| 4.3.3. Resultados y discusión..... | 123 |
| 4.4. Artículo 4:..... | 125 |
| Structural analysis of glucuronoxylan-specific Xyn30D and its attached CBM35 domain gives insights into the role of modularity in specificity | 125 |
| 4.4.1. Introducción al análisis estructural de la glucuronoxilanasa Xyn30D y de su módulo auxiliar CBM35 como determinante de la especificidad..... | 127 |
| 4.4.2. Publicación..... | 129 |
| 4.4.3. Resultados | 144 |
| 4.4.4. Discusión | 146 |
| 4.4.5. Conclusiones: Artículos 3 y 4 | 148 |
| 5. Discusión General | 151 |
| 6. Conclusiones | 159 |
| 7. Bibliografía | 163 |

| | |
|---|------------|
| 8. Anexos | 171 |
| Anexo I - Tabla periódica de los elementos | 173 |
| Anexo II - Tabla de aminoácidos | 175 |
| Anexo III - Simetría cristalina | 177 |

1. INTRODUCCIÓN

“Lo último que uno sabe acerca de una tarea es por dónde empezar”

Blaise Pascal (Francia, 1623- 1662)

Filósofo, físico y matemático

1. Introducción

Las enzimas son potentes herramientas que pueden ser usadas para la protección del medioambiente de muy diversas formas. Su presencia en la industria está creciendo de manera muy significativa y cada vez se emplean más en empresas petroleras, agroalimentarias, textiles, papeleras, químicas, bioquímicas y farmacéuticas. Además, también ayudan a mantener un ambiente libre de contaminación a través de su uso en la gestión de residuos. La tecnología recombinante del ácido desoxirribonucleico (ADN), la ingeniería de proteínas y el diseño racional de enzimas son áreas de investigación emergentes relacionadas con las aplicaciones ambientales de las enzimas.

En este contexto, se ha desarrollado la tesis doctoral que a continuación se presenta, la cual se ha enfocado hacia el estudio estructural de dos xilanasas microbianas de *Paenibacillus barcinonensis* con importantes aplicaciones biotecnológicas. Estas enzimas catalizan la hidrólisis de uno de los polisacáridos más abundantes en la naturaleza como es el xilano, el cual se encuentra en un alto porcentaje dentro de la pared celular vegetal. Sumando su poder biocatalizador al cada vez más creciente empleo de materiales vegetales en la industria, las xilanasas han adquirido un gran atractivo comercial. Biotecnológicamente hablando, su utilidad frente a catalizadores químicos y procesos industriales convencionales radica en que son biocatalizadores altamente selectivos, con elevada eficiencia y bajo impacto ambiental, requisitos indispensables para el desarrollo de tecnologías verdes (Himmel & Bayer, 2009; Sticklen, 2008).

1.1. Pared celular vegetal e importancia fisiológica

El término *cell* (celda, célula) fue empleado por primera vez por Robert Hooke para describir los huecos o celdas que apreció al observar una lámina de corcho en su microscopio óptico elemental (Hooke, 1665). Lo que verdaderamente estaba viendo eran las paredes celulares definidas poligonalmente.

La pared celular se encuentra en todas las células vegetales excepto en las esporas móviles de algas y hongos y en las células sexuales de las embriofitas^a. Se localiza en la zona exterior de la célula, en unión íntima con la membrana plasmática (Figura 1.1). Es el exoesqueleto que mantiene en equilibrio la presión osmótica, ofrece soporte mecánico y protege a la célula. Tiene un espesor variable que aumenta según el grado de diferenciación de la célula y depende también de la función del tejido. En el caso de xilema^b y esclerénquima^c, la

^a Embriofita: adj. y n. f. Bot. Se aplica a los vegetales que tras la fecundación del óvulo desarrollan un embrión.

^b Xilema: n. m. Bot. Tejido leñoso de las plantas vasculares que transporta principalmente agua y minerales de una parte a otra de la planta.

^c Esclerénquima: n. m. Bot. Tejido constituido por células de membrana totalmente engrosada y lignificada, propio de los órganos en estado de desarrollo completo.

pared adquiere importancia comercial dado que es la que forma la madera y algunas fibras textiles. Su composición química y grosor influye en la calidad de ciertos productos alimenticios como las frutas, los tubérculos o el maíz.

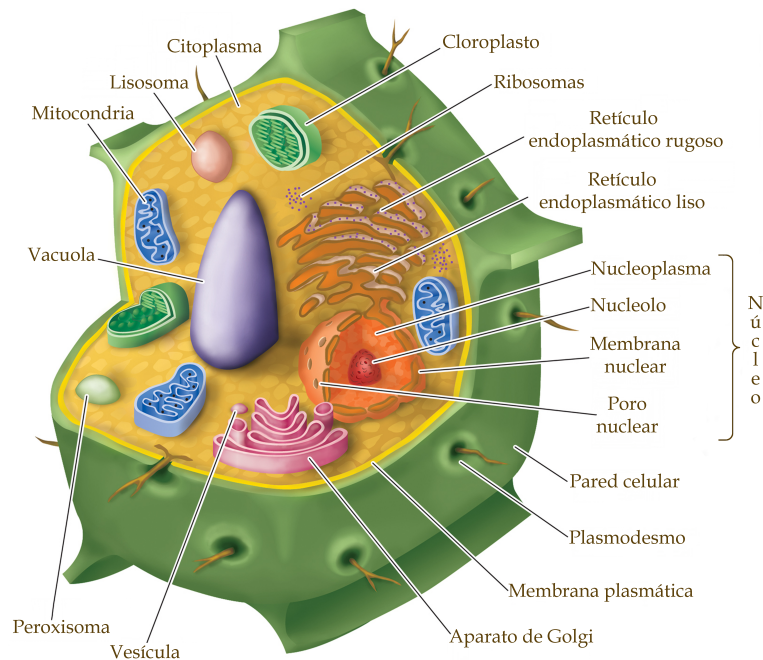


Figura 1.1. Esquema de la pared celular vegetal. Figura adaptada de la Enciclopedia Británica^d.

La pared celular vegetal tiene un papel de gran importancia no sólo en las células vivas, también en células muertas. En las primeras, es esencial en funciones vitales como absorción, transpiración, secreción, defensa o reconocimiento, entre otras. En las segundas, también tienen funcionalidad ya que, por ejemplo en los árboles, la mayor parte de la corteza y la madera están formadas únicamente por paredes celulares. En la corteza, presentan materiales que protegen las células de la desecación, y en la madera, sirven como soporte mecánico ya que son gruesas y rígidas. Ejercen, además, una barrera estructural frente a macromoléculas y a patógenos. Debido a que la pared celular vegetal se conserva bien, se puede realizar el estudio de la misma fácilmente en plantas secas y fósiles.

En procariotas, hongos, algas y plantas, la pared celular es esencial en funciones primarias como determinación de la forma y regulación del volumen celular. Sin embargo, a diferencia del resto, la pared celular vegetal tiene funciones adicionales que hacen que su estructura y composición sea mucho más diversa y compleja. Además, la mayoría del carbono asimilado en la fotosíntesis finaliza como polisacárido en la pared celular. Estos polímeros pueden ser hidrolizados y usados para la generación de nuevos compuestos. Los oligosacáridos presentes en la pared actúan como importantes moléculas de señalización durante la diferenciación celular y el reconocimiento de patógenos

^d Cytoplasm. (2015). En *Encyclopædia Britannica*. Recuperado de <<http://www.britannica.com/science/cytoplasm>>.

y simbiontes. Asimismo, ejerce de barrera a la difusión de macromoléculas y patógenos.

Por último, y además de su papel biológico esencial, la pared celular vegetal presenta un marcado interés económico, ya que se emplea como producto natural en múltiples industrias como textil, papelera, maderera, petrolera o biotecnológica. Del mismo modo, al ser una importante fuente de fibra, juega un papel importante en la nutrición humana (Apartado 1.5).

1.1.1. Composición química y estructura

Las paredes celulares de las plantas son amalgamas increíblemente complejas de hidratos de carbono, proteínas, lignina, cutina^e, suberina^f, y ciertos compuestos inorgánicos que varían según la especie vegetal, tipo de célula e, incluso, células adyacentes (Showalter, 1993). Se denomina lignocelulosa al polímero formado por celulosa y hemicelulosa que está unido fuertemente a la lignina, el cual es responsable de la rigidez de las paredes celulares vegetales y que supone una de las mayores fuentes de carbono renovables para la producción de biocombustibles.

La estructura, conformación, distribución y composición de la pared celular varía en función de los taxones, el tejido, la edad y el tipo de célula, así como dentro de cada capa de la misma (Ding & Himmel, 2006; Sticklen, 2008). Además, son compartimentos dinámicos que cambian a lo largo de la vida de la célula. Morfológicamente hablando, se pueden diferenciar tres zonas por microscopía electrónica (Figura 1.2): lámina media, pared primaria y pared secundaria. La pared es secretada por la célula viva, de manera que la capa más antigua está hacia el exterior siendo la más joven la interna.

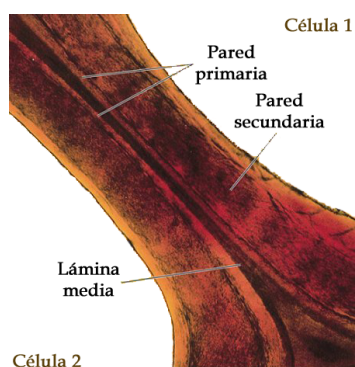


Figura 1.2. Imagen de la pared celular vegetal y sus partes. Figura adaptada del libro *The Living World* 5ª Edición.

Los componentes principales de la pared celular vegetal se representan de forma esquemática en la Figura 1.3 y se describen a continuación:

^e Cutina: n. f. Bot. Sustancia impermeable de naturaleza lipídica localizada en la parte externa de algunas células epidérmicas y que procede de la transformación de la celulosa en el momento de la diferenciación de los tejidos.

^f Suberina: n. f. Bot. Polímero natural producido por las paredes celulares de algunas células de las plantas. Está formada por ácidos grasos hidróxidos y epóxidos unidos por enlaces éster.

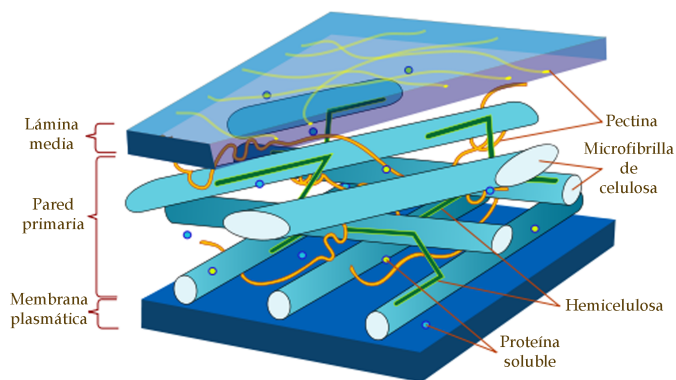


Figura 1.3. Componentes principales de la pared celular vegetal. Figura adaptada de <http://www.wpclipart.com>.

- **Lámina media:** Es la primera capa que se deposita y se inicia en el momento de la división celular (Figuras 1.2 y 1.3). Está formada fundamentalmente por sustancias pécticas (ácido péctico: unión de moléculas de ácido galacturónico^g; pectatos: ácido péctico e iones metálicos, Ca^{2+} o Mg^{2+}). Es amorfa y ópticamente inactiva siendo complicada su observación en el microscopio óptico (Heredia *et al.*, 1995). Así, para dos células vecinas, las paredes primarias y la lámina media presentes entre ambas forman una unidad que con el microscopio óptico se llega a apreciar como un todo y se denomina lámina media compuesta.
- **Pared primaria:** Una vez formada la lámina media, los carbohidratos se van depositando a ambos lados de la misma, constituyendo la pared primaria de cada una de las células contiguas (Figuras 1.2 y 1.3). Está presente en todas las células vegetales y controla el crecimiento celular, siendo la base estructural del esqueleto de la planta. Se forma después de la división celular y antes de la finalización del crecimiento de la célula, localizándose en células jóvenes. Suele tener poco espesor y una arquitectura molecular similar en casi todos los tipos celulares. Está constituida por celulosa, hemicelulosa, pectina y, en menor cantidad, glicoproteínas estructurales unidas entre sí mediante enlaces covalentes. Existen dos tipos de pared celular primaria que se diferencian según el tipo de entrecruzamiento. Las **paredes tipo I** están presentes en plantas dicotiledóneas^h presentando igual cantidad de glucano y xiloglucano (XG) impregnados en una matriz de pectina. Las **paredes tipo II** se encuentran en las plantas herbáceas y tienen glucuronoarabinosilanos (GAX) como ramificación del glucano careciendo de pectina y proteínas estructurales (Sticklen, 2008). Concretamente en las endospermas de cereales, glucanos β -(1,3-1,4) mixtos y arabinosilanos (AX) son los componentes principales de la pared celular.

^g Ácido D-galacturónico (GlaA): Quím. org. Es un monosacárido de 6 átomos de carbono correspondiente a la forma oxidada de la D-galactosa.

^h Dicotiledóneas: f. pl. Bot. Clase del subtipo de las angiospermas constituida por plantas que tienen dos cotiledones en su embrión, por ejemplo la judía y la malva.

- **Pared secundaria:** Se forma sobre las paredes primarias hacia el interior de la célula cuando se ha detenido su crecimiento y elongación (Figuras 1.2 y 1.4). En este momento, existen células especializadas que pueden seguir sintetizando polímeros hasta desarrollar la pared secundaria, la cual está muy especializada en composición y arquitectura, marcando el grado de diferenciación de la célula. Es considerada una pared supletoria con una función predominantemente mecánica. La gran diferencia con las anteriores capas es la lignificación, proceso que ocurre en células relacionadas con una función de soporte. Esta capa presenta más cantidad de celulosa, una menor cantidad de hemicelulosa y, en algunas clases celulares, lignina y/o ceras. Habitualmente, está formada por tres capas denominadas (del exterior al interior), S1, S2 y S3 (Figura 1.4). Su estructura es más gruesa proporcionando mayor resistencia y rigidez estando presente, por ejemplo, en los troncos de árboles.

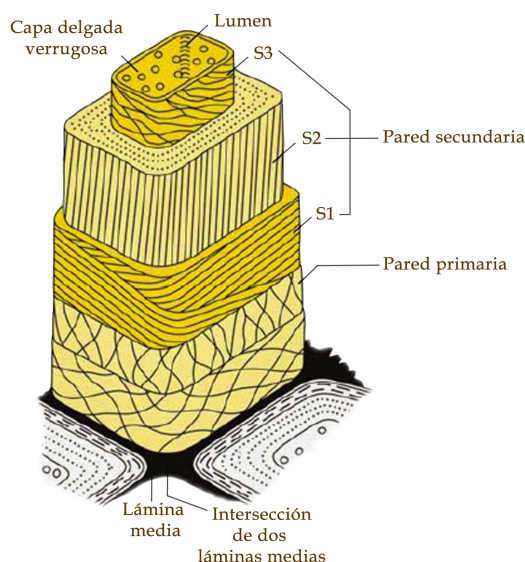


Figura 1.4. Estructura de la pared con sus diferentes capas. La pared secundaria tiene tres secciones diferentes con microfibrillas de celulosa en distintas orientaciones (capas S1, S2 y S3). Figura adaptada de www.citharaworld.blogspot.com.

En resumen, la composición y estructura de la pared celular vegetal depende del tipo celular, especie y tejido. Puede decirse que está compuesta principalmente por **celulosa**, **hemicelulosa**, **sustancias pécticas**, **glicoproteínas** y, en ocasiones, **lignina**.

1.1.2. Celulosa

Las plantas producen alrededor de 180 billones de toneladas de celulosa al año por lo que es la mayor reserva de carbono orgánico en la Tierra (Festucci-Buselli *et al.*, 2007). Constituye un 15-30 % de la masa seca de la pared primaria y hasta un 40 % de la secundaria (Sticklen, 2008) presentando un papel fundamentalmente estructural ya que forma parte de los tejidos de soporte. Es

un homopolímero lineal compuesto por moléculas de glucosa enlazadas mediante enlaces glucosídicos β -(1,4) donde cada unidad de D-glucosa está girada 180° respecto a la contigua a lo largo del eje molecular, por lo que la unidad repetitiva estructural es la formada por dos residuos de D-glucosa: la celobiosa (Delmer & Amor, 1995). La celulosa tiene una sofisticada organización y su estructura básica es la **fibrilla elemental**, la cual está compuesta por unas 100 moléculas de celobiosa. Al adherirse alrededor de 20 de estas fibrillas elementales, se constituye lo que se denomina **microfibrilla**, la cual se puede observar con el microscopio electrónico y tiene un tamaño de 15-20 nm. Por unión de casi 250 microfibrillas, se forma la **macrofibrilla**, en este caso observable al microscopio óptico. Finalmente, la asociación de aproximadamente 1 500 microfibrillas constituye la fibra de celulosa, visible con la lupa, parte esencial de su fase cristalina (Figura 1.5). Esta molécula puede presentar un grado de polimerización (DPⁱ) mayor de 15 000 (Heredia *et al.*, 1995). Estudios estructurales de rayos X sugieren que su estructura está rodeada por otra menos organizada en la que otros polisacáridos (xilanos y mananos) están unidos a la celulosa mediante enlaces covalentes.

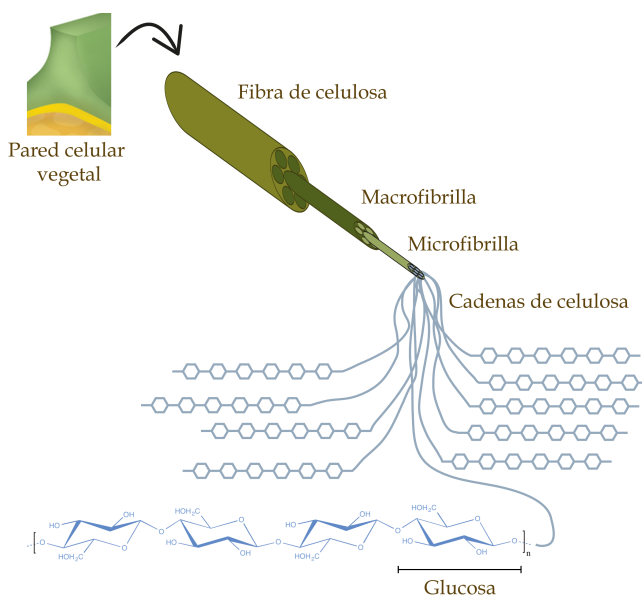


Figura 1.5. Microfibrillas de celulosa en la pared celular vegetal.

La estructura que puede presentar la microfibrilla es muy diversa. Mientras que en la pared primaria las fibrillas se encuentran entrelazadas aparentemente al azar (Figura 1.6a), en la pared secundaria se observan con una disposición ordenada (Figura 1.6b). La disociación de la pared secundaria en capas resulta de la orientación de las fibrillas: en una casi horizontal, en la siguiente casi vertical, y en la tercera nuevamente casi horizontal (Figura 1.4). En el algodón, por ejemplo, la mayor parte de esta pared secundaria presenta microfibrillas orientadas de forma helicoidal mientras que en el lino, se dispone una capa frente a otra de forma opuesta.

ⁱ Grado de polimerización (DP): Es el número de veces que se repite la unidad monomérica en una cadena. Se calcula dividiendo el correspondiente promedio del peso molecular por el peso de la unidad monomérica.

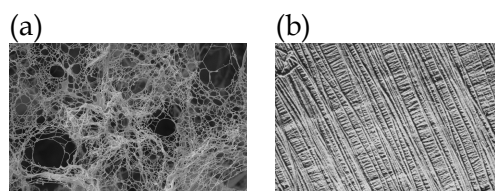


Figura 1.6. (a) Pared primaria: Las fibrillas se encuentran entrelazadas al azar. (b) Pared secundaria: Las fibrillas se observan con una disposición ordenada. Figuras adaptadas de (Tingaut *et al.*, 2011) y www.nzetc.victoria.ac.nz.

1.1.3. Hemicelulosa

Es la segunda estructura orgánica más abundante en la pared celular vegetal. El término “hemicelulosa” fue introducido por primera vez por Schulze (Motta *et al.*, 2013) para nombrar a aquellas fracciones extraídas del material vegetal mediante tratamiento alcalino. Son polisacáridos heterogéneos que presentan una cadena principal con ramificaciones cortas. La composición varía de un tipo de célula a otra y según el tipo de especie. Los cuatro tipos de polímeros más habituales que suelen estar presentes son: xiloglucanos, glucanos mixtos, xilanos y mananos.

Aunque se puede establecer una clasificación de los polisacáridos en función de su extracción con tratamiento alcalino, esta sería insuficiente e incompleta debido a la gran diversidad estructural y fisicoquímica que presentan estos polisacáridos. Una de las clasificaciones que se puede hacer para la hemicelulosa es agruparla en xiloglucano, heteroxilano, heteromanano y β -(1,3-1,4)-glucanos. Forman parte de un grupo muy diverso y todos tienen en común que presentan ramificaciones en su cadena. Recientemente se ha aceptado que el término hemicelulosa, se emplee para polisacáridos únicamente con la configuración de C1 y C4 en posición ecuatorial (Scheller & Ulvskov, 2010).

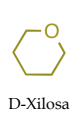
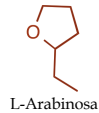
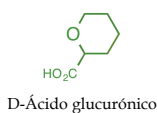
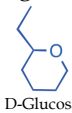
La estructura y distribución de hemicelulosa en la pared primaria y secundaria de las plantas se representa en la Tabla 1.1:

Tabla 1.1. Ilustración esquemática de los tipos de hemicelulosas encontradas en las paredes celulares vegetales.

| Polisacárido (Abreviatura) | Fórmula |
|----------------------------|---|
| Xiloglucano (XG) | <p>Cadenas de glucosa β-(1,4) con cadenas laterales de xilosa α-(1,6). Puede presentar ramificaciones adicionales como D-galactosa (-R).</p> |
| Heteroxilano | <p>Cadenas de xilosa β-(1,4) con ramificaciones laterales cortas de otros azúcares y/o grupos acetilos en posición C2 y/o C3 de las xilosas. Se pueden clasificar en varios tipos según las ramificaciones que presenten.</p> <p>Glucuronoarabinoxilano (GAX)</p> <p>Glucuronoxilano (GX)</p> |
| Heteromanano | <p>Polímero de manosa β-(1,4).</p> <p>Cadena de manosas con sustituciones de galactosas.</p> <p>Galactomananos (GM)</p> <p>Cadena de manosa y glucosa con ramificaciones de galactosa.</p> <p>Galactoglucomananos (GGM)</p> <p>Mezcla β-glucanos 1,3-1,4. Cada 3 o 4 residuos.</p> <p>β-(1,3-1,4)-glucanos (GS)</p> |

“Fer” representa la esterificación con ácido ferúlico (ácido 3-metoxi-4-hidroxicinámico).

Código de colores:



Las principales diferencias existentes entre celulosa y hemicelulosa se recogen en la Tabla 1.2 (Pereira *et al.*, 2008):

Tabla 1.2. Diferencias entre celulosa y hemicelulosa.

| Celulosa | Hemicelulosa |
|--|---|
| Unidades de glucosas | Unidades de pentosas y hexosas |
| Alto DP (2 000-18 000) | Bajo DP (50-300) |
| Presenta disposiciones fibrosas | No presenta disposiciones fibrosas |
| Presenta fracción cristalina y amorfa | Sólo tiene fracción amorfa |
| Lentamente atacado por ácidos inorgánicos diluidos en soluciones calientes | Rápidamente atacado por ácidos inorgánicos diluidos en soluciones calientes |
| Insoluble en álcalis | Soluble en álcalis |

1.1.4. Pectina

Las pectinas son polímeros altamente heterogéneos tradicionalmente caracterizados por ser relativamente fácil su extracción con ácidos en caliente o productos quelantes y poseer gran cantidad de residuos de ácido galacturónico (GlaA) (Figura 1.7a). En cooperación con la hemicelulosa, se agregan a las microfibrillas de celulosa y promueven la hidratación gracias a su elevada capacidad hidrofílica. De esta forma se crea una fase tipo gel en la cual están sumergidas celulosa y hemicelulosa, regulando la porosidad de la pared celular. Este tipo de polímeros tienen dos regiones diferenciadas: lisa/regular y rugosa/complicada. La primera está constituida por residuos de ácido α -1,4-D-galacturónico, los cuales pueden estar acetilados en las posiciones O2/O3 o metilados en O6. En la segunda región, se pueden identificar dos estructuras diferentes (de Vries & Visser, 2001):

- Xilogalacturonano: Ramificaciones de D-xilosa en la cadena galacturonano^j.
- Ramnogalacturonano I: A la cadena de galacturonano se le alternan residuos de L-ramnosa (Figura 1.7b) mediante enlaces α -(1,2). Además, se pueden unir grandes cadenas de galactano^k y arabinano^l al O4 de la ramnosa.

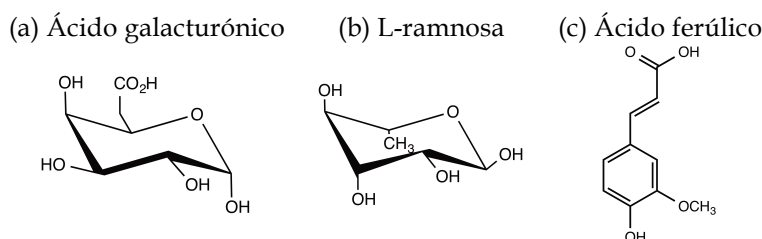


Figura 1.7. Fórmulas de los principales componentes de la pectina.

^j Galacturonano: Cadenas de GlaA α -(1,4), metiladas en el C6 de los grupos carboxilos en diferente proporción.

^k Galactano: Cadenas de D-galactosa β -(1,4) que presentan en ocasiones residuos de ácido ferúlico (Figura 1.7c) unidos al O6 de alguna de ellas.

^l Arabinano: Cadenas de L-arabinosa α -(1,5) que pueden tener residuos de L-arabinosa unidos por enlaces α -(1,3) a la cadena principal y/o residuos de ácido ferúlico unidos al O2 de la arabinosa terminal.

1.1.5. Lignina

Después de la celulosa y el xilano, es uno de los polímeros más abundantes en la tierra. Tiene una gran resistencia a la degradación química y biológica.

Cuando se ha terminado la fase de crecimiento del tejido de la planta, algunas células diferenciadas pueden llegar a sintetizar lignina, la cual se puede describir como un polímero formado por unidades de fenilpropano. Así, la lignina se forma por la polimerización de tres alcoholes aromáticos con la estructura del fenilpropano: *p*-cumarílico, coniferílico y sinapílico (Figura 1.8a, b, c), unidos entre sí por enlaces carbono-carbono o enlaces tipo éter entre los anillos aromáticos mediante deshidrogenación y polimerización (Chávez-Sifontes & Domine, 2013; Heredia *et al.*, 1995).

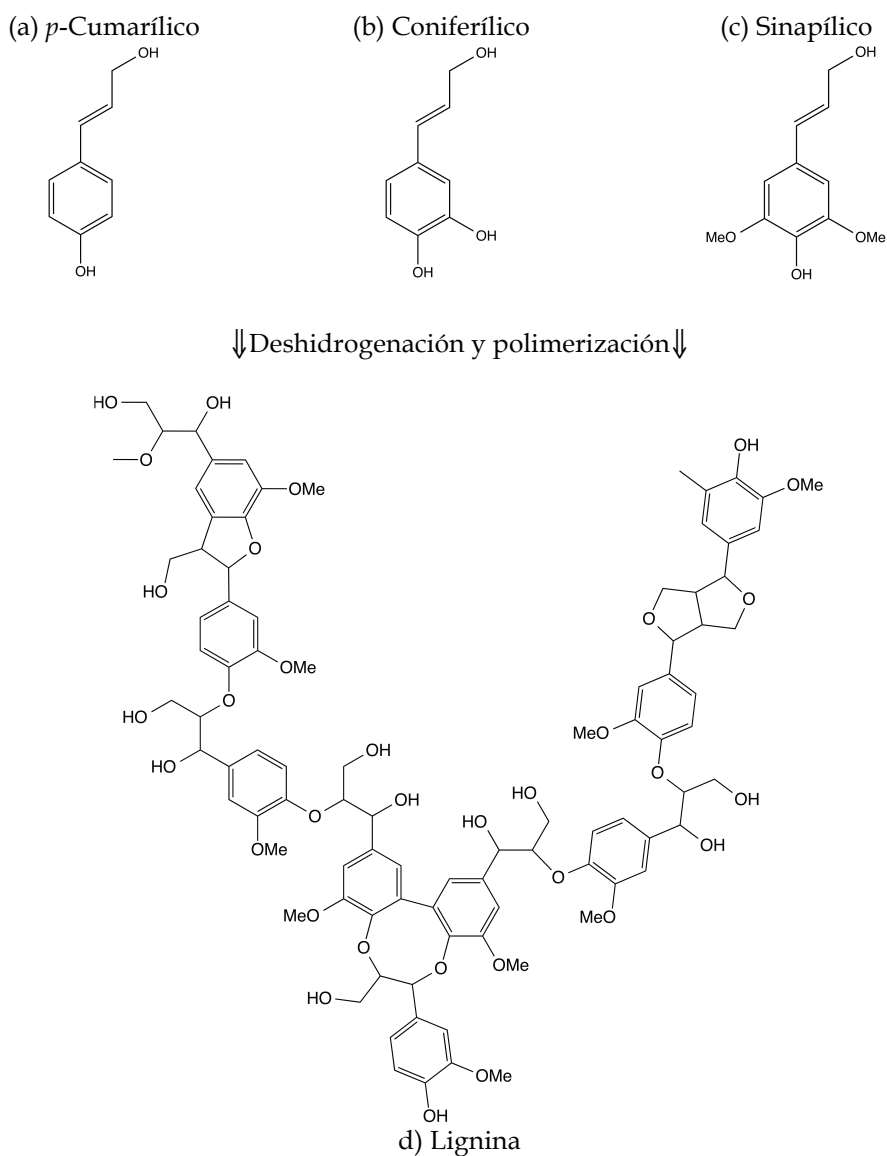


Figura 1.8. (a, b, c) Fórmulas de los precursores de la lignina. (d) Fórmula de la lignina.

Lignina y hemicelulosa forman una matriz amorfa dentro de la cual están las fibras de celulosa. Es el componente más resistente del material lignocelulósico por lo que dificulta no sólo su propia hidrólisis si no también la de los polisacáridos que están en contacto con ella. Constituye del 20-30 % de la madera de los árboles. Dentro de las maderas blandas, predomina el alcohol coniferílico mientras que en maderas duras se encuentra el alcohol cumárico y el sinapílico (Quiroz-Castañeda & Folch-Mallol, 2011).

La función principal de la hemicelulosa es la conexión entre las fibras de celulosa y la lignina. De esta forma, proporciona una gran rigidez al entramado celulosa/hemicelulosa/lignina. Hemicelulosa y lignina están unidas esencialmente mediante enlaces de tipo éster entre la arabinosa de la hemicelulosa y los grupos hidroxilos de la lignina, mientras los puentes de hidrógeno son los responsables de la unión con la celulosa (Laureano-Perez *et al.*, 2005).

1.2. El xilano

El xilano es el componente mayoritario de la hemicelulosa, constituyendo más del 70 % de su masa (Quiroz-Castañeda & Folch-Mallol, 2011). Después de la celulosa, es el polisacárido renovable más abundante en la naturaleza (Biely, 1985). Se encuentra sobre todo en las paredes secundarias, siendo el componente principal de las paredes celulares del tejido leñoso maduro. Por otra parte, los xilanos son componentes esenciales de la hemicelulosa en paredes primarias de monocotiledóneas (Wong *et al.*, 1988). Un ejemplo que refleja la importancia del xilano como reserva de carbono está representado por la madera de abedul, ya que el 35 % de su peso seco es xilano (Timell & Syracuse, 1967). En general, el xilano es el componente principal de la hemicelulosa de la madera en angiospermas, pero es menos abundante en la de gimnospermas. Así, representa del 15-30 % y 7-12 % del peso total en seco, respectivamente (Wong *et al.*, 1988).

Como se ha expuesto en el apartado anterior, este polímero está constituido por una cadena principal formada por monómeros de D-xilosa unidos mediante enlaces β -1,4 (Figura 1.9).

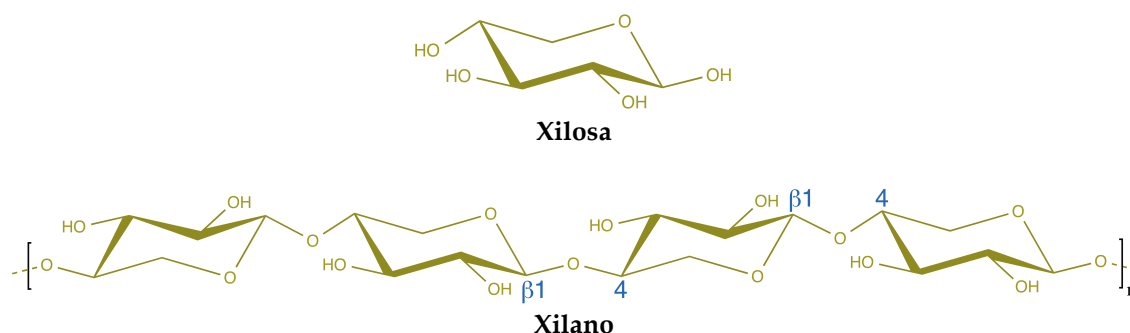


Figura 1.9. Fórmulas de la D-xilosa y el xilano.

Únicamente se ha caracterizado un xilano casi puro: el esparto (*Stipa tenacissima*), el cual contiene 95-98 % de xilosa (Heredia *et al.*, 1995). Sin embargo, lo más habitual es encontrar xilanos ramificados. Los grupos secundarios más comunes que suele presentar este polisacárido, en las posiciones C2 y C3, pueden ser: L-arabinosa, ácido glucurónico (GlcA), ácido metilglucurónico (MeGlcA), D-galactosa, grupo acetil, ácido ferúlico (formando puentes diferúlico entre cadenas de xilanos) y/o ácido *p*-cumárico (Figura 1.10) (de Vries & Visser, 2001).

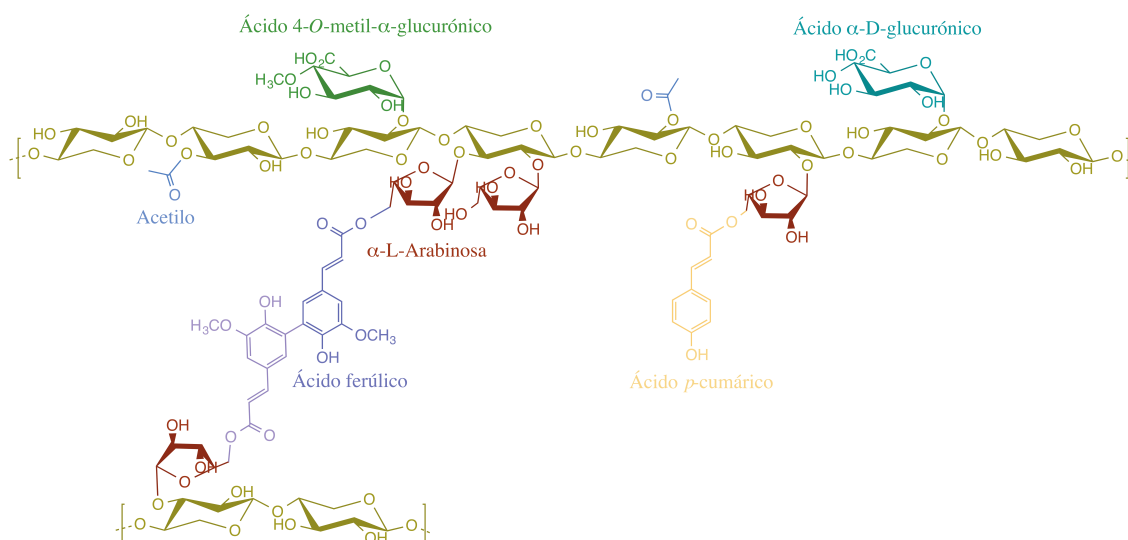


Figura 1.10. Esquema de la estructura de una molécula de xilano y sus ramificaciones.

Los xilanos de muchos tipos de plantas se obtienen como heteropolisacáridos y los términos **GX** (glucuronoxilano) y **GAX** son comúnmente usados para describir xilanos de maderas duras y blandas^m, respectivamente. Estos dos tipos de xilano tienen residuos de ácido 4-O-metil- α -D-glucurónico unidos al C2 de las unidades de xilosa. Las maderas duras tienen esta sustitución en aproximadamente el 20 % de los residuos de xilosa, mientras que las maderas blandas lo presentan en alrededor del 10 %. Así, los xilanos de la madera, con un gran porcentaje de ácido D-glucurónico como ramificación, son conocidos como GX (de Vries & Visser, 2001). Las maderas blandas pueden presentar también ramificaciones de α -L-arabinosa en el C3 en aproximadamente el 13 % de los residuos de xilosa (Coughlan & Hazlewood, 1993). De este modo, los xilanos de cereales contienen grandes cantidades de L-arabinosa y se les denomina arabinoxilanos o más globalmente GAX.

El DP es muy variable en los xilanos y puede llegar a ser del orden de 150-200 en maderas duras en contraste con 70-130 en maderas blandas (Kulkarni *et al.*, 1999).

^m Madera dura: Madera procedente de un árbol de hoja caduca y ancha, perteneciente a las angiospermas, como el cerezo, el roble, el arce, álamo, abedul, olmo, etc. Madera blanda: Madera procedente de las coníferas (pino, cedro) que es relativamente blanda y fácil de aserrar y trabajar.

Por otra parte, el grado de acetilación que presenta el xilano de maderas duras determina su solubilidad (Biely, 1985), pudiendo contener grupos acetilos en las posiciones C2 y C3 de las cadenas de xilosa hasta en un 70 %. Las plantas herbáceas presentan una alta proporción de arabinosa en posición C2 y/o C3. Por último, los ácidos ferúlico y cumárico pueden estar enlazados a la cadena de xilosas a través de su unión a la arabinosa en posición C5. Se han encontrado en diferentes tipos de xilanos y su existencia está asociada a enlaces covalentes reticulares entre moléculas de xilano o con la lignina (Wong *et al.*, 1988).

1.2.1. Degradación enzimática del xilano

La hemicelulosa tiene un alto potencial industrial, ya que los productos que generan su hidrólisis o conversión tienen un gran valor añadido (Figura 1.11). La biodegradación de la hemicelulosa produce azúcares monoméricos y ácido acético. La degradación del xilano, como principal carbohidrato de ésta, es de suma importancia en este proceso.

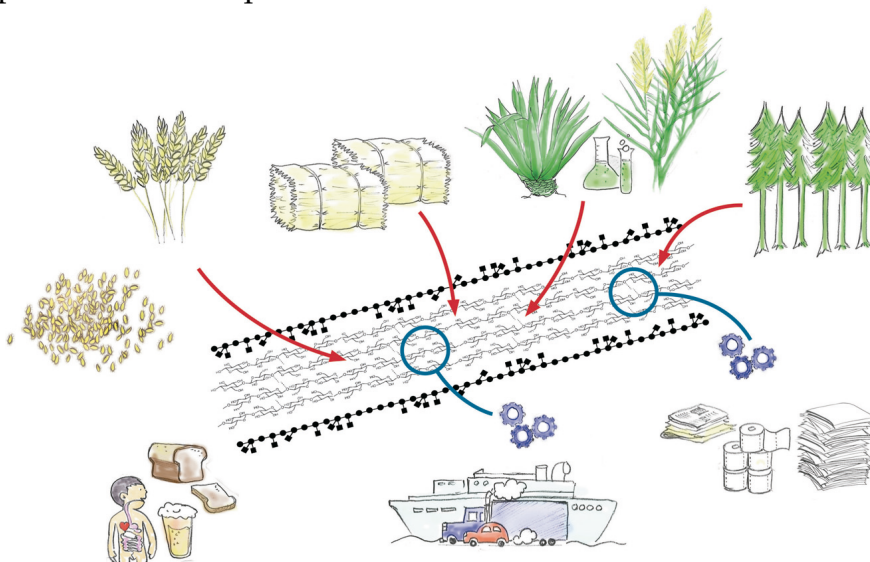


Figura 1.11. Representación general de las aplicaciones biotecnológicas que puede tener la degradación enzimática del xilano. Imagen adaptada de (Burton & Fincher, 2014). Cortesía de Bruno Carrocci.

Debido a la naturaleza compleja y heterogénea del xilano, la completa degradación de su esqueleto requiere de la acción cooperada y coordinada de una gran variedad de enzimas hidrolíticas (Biely, 1985; Coughlan & Hazlewood, 1993). Se puede hacer una clasificación general de estas enzimas en dos tipos según la ruptura del xilano sea en su cadena principal o en las cadenas laterales. Dentro de estas, las xilanasas (endo-1,4- β -xilanasas; ECⁿ 3.2.1.8) se encargan de la escisión de la cadena principal del polímero y las β -xilosidasas (1,4- β -xilosidasas; EC 3.2.1.37) fragmentan las unidades de xilosas de xilobiosas y xilooligómeros terminales. Por su parte, la degradación de las cadenas laterales del xilano está catalizada por α -L-arabinofuranosidasas (EC

ⁿ EC: Número de la Comisión de Enzimas (*Enzyme Commission*).

3.2.1.55), α -D-glucuronidasas (EC 3.2.1.139), acetilxilano esterasas (EC 3.1.1.72), feruloil esterasas (EC 3.1.1.73) y *p*-cumárico esterasas (Figura 1.12).

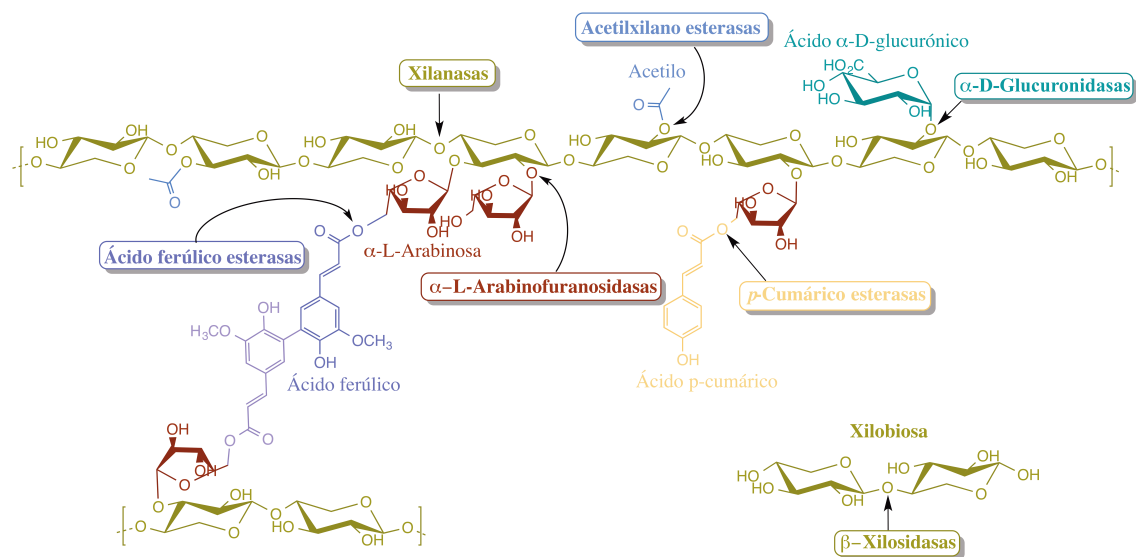


Figura 1.12. Esquema de las enzimas implicadas en la degradación del xilano y enlaces sobre los que actúan.

1.3. Xilanasas y su mecanismo catalítico

Como se ha destacado en el apartado anterior, las xilanasas (endo-1,4-β-xilanasas) son enzimas clave en la degradación de la hemicelulosa. Muchos microorganismos pueden producir múltiples xilanasas (Collins *et al.*, 2005). Estos pueden tener diversas estructuras, propiedades fisicoquímicas, actividades específicas y rendimientos. Esta variedad puede aumentar la eficiencia en la hidrólisis, pero también incrementa la diversidad y complejidad en las enzimas. Dos ejemplos típicos de microorganismos que producen isoenzimas ° xilanolíticas son *Aspergillus niger* (secreta 16 xilanasas) y *Trichoderma viride* (produce 13 xilanasas) (Biely, 1985).

Las xilanasas son enzimas tipo “endo” que hidrolizan enlaces internos dentro del xilano, dando una mezcla de xilooligosacáridos (XOS) procedentes del polímero (Figura 1.13). No obstante, la caracterización de xilanasas de *Aeromonas* ha sugerido un modo de hidrólisis diferente debido a que genera un producto de reacción único, lo cual es la consecuencia de una ruptura tipo “exo” (Kubata *et al.*, 1995). Sin embargo, como este tipo de mecanismo requiere primero de la quiebra de las cadenas laterales del xilano, la probabilidad de que únicamente exoxilanasas puras produzcan la degradación completa del mismo parece improbable.

° Isoenzimas: Dos o más enzimas que son químicamente distintas pero que funcionan de forma similar.

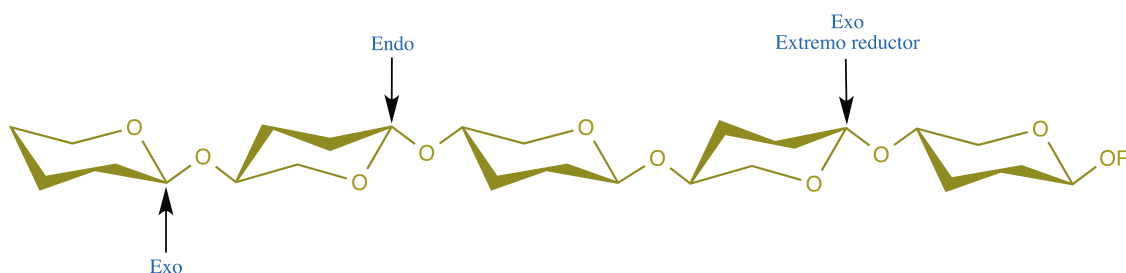


Figura 1.13. Esquema del polímero de xilano indicando los tipos de ruptura: “endo” y “exo”.

Mecanismo catalítico de las xilanasas

La hidrólisis enzimática del enlace glicosídico está catalizada por distintas enzimas denominadas glicosil hidrolasas (GH), dentro de las cuales se encuentran las xilanasas. Esta reacción tiene lugar a través de la intervención de dos aminoácidos presentes en el sitio catalítico, uno de los cuales actúa como ácido o electrófilo (donante de protones) y otro como base o nucleófilo (que cede electrones). En la mayoría de los casos son residuos carboxílicos, generalmente un aspartato o un glutamato (Davies & Henrissat, 1995). Se han descrito dos mecanismos de hidrólisis en función de la configuración final del carbono anomérico. Una primera opción es que su configuración se mantenga. Así la reacción de hidrólisis tendría lugar a través de un mecanismo de retención (Figura 1.14a). La segunda opción es que la configuración anomérica cambie. De esta forma la reacción de hidrólisis se llevaría a cabo a través de un mecanismo de inversión (Figura 1.14b). Así, dependiendo de la posición espacial de estos aminoácidos, la reacción tiene lugar a través de uno u otro mecanismo. Si estos residuos carboxílicos están dispuestos a una distancia de aproximadamente 5,5 Å, se da el mecanismo de retención y si, por el contrario, la distancia es de 9-10 Å, ocurre el mecanismo de inversión.

La hidrólisis a través del mecanismo de retención se da en dos etapas. En la primera, tiene lugar el ataque del residuo nucleófilo al centro anomérico, dando como resultado la formación de un complejo intermedio enzima-glicósido. En la segunda etapa el carboxilato, que inicialmente actuó como ácido de la reacción protonando el oxígeno glicosídico, actúa ahora como base captando el protón de la molécula de agua, rompiendo este estado intermedio y regenerando la enzima (Zechel & Withers, 2001). Por otro lado, en el mecanismo de inversión la hidrólisis se realiza en una sola etapa ya que como consecuencia de la mayor distancia entre los residuos catalíticos, la molécula de agua puede alojarse entre la base y el azúcar (Davies & Henrissat, 1995).

En 1997, Davies formuló la nomenclatura según la cuál las posiciones del sustrato a hidrolizar se denominan de $-n$ a $+n$, de manera que $-n$ representa el extremo no-reductor y n el extremo reductor, ocurriendo la ruptura del enlace entre el residuo -1 y $+1$ del azúcar (Apartado 1.3.2) (Davies *et al.*, 1997).

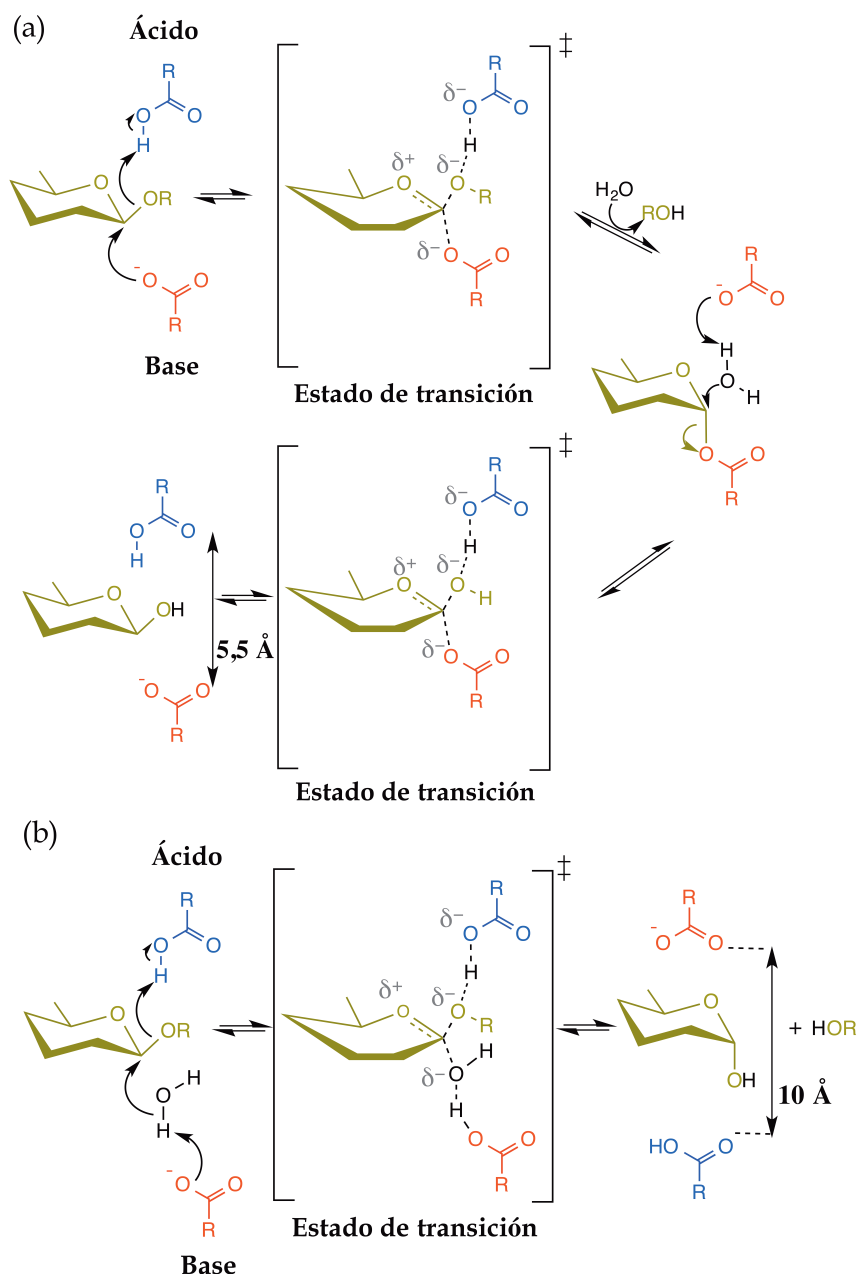


Figura 1.14. Mecanismo general de hidrólisis. (a) Con retención. (b) Con inversión.

Como el xilano es un polímero largo que no puede penetrar dentro de las células, las xilanasas en general tienen que ser secretadas a un ambiente extracelular para poder realizar su función. Aunque la mayoría de las xilanasas son enzimas extracelulares, generalmente secretadas mediante el camino sec-dependiente (Tjalsma *et al.*, 2004), han sido descritas xilanasas localizadas en el periplasma de bacterias Gram negativas, como es el caso de la bacteria del rumen *Probotella bryantii* (el 80 % de la actividad xilanasas se localiza en el periplasma) (Miyazaki *et al.*, 1997), de *Cellvibrio mixtus* (Fontes *et al.*, 2000) y precisamente de una xilanasas de *P. barcinonensis*, caracterizada en 2003, Xyn10B, que resultó ser intracelular (Gallardo *et al.*, 2003). Todos estos autores sugieren que la función de estas xilanasas no sería la degradación del xilano sino que,

muy probablemente, sus sustratos sean XOS de menor tamaño que puedan atravesar la membrana externa. La ventaja funcional sería que, de este modo, estas enzimas quedarían protegidas de la acción de proteasas extracelulares.

Generalmente, la inducción de xilanasas se produce en microorganismos que durante su crecimiento están en contacto con sustratos que tienen xilano. Así, la producción de un sistema multienzimático de xilanasas, en el cual cada enzima tiene una función específica, representa una estrategia para lograr una hidrólisis eficiente del xilano.

Las xilanasas también pueden estar presentes en algunos microorganismos anaeróbicos dentro de estructuras extracelulares complejas denominadas celulosomas (Figura 1.15). Estos son complejos multienzimáticos cuyos componentes actúan de manera sinérgica para hidrolizar la celulosa (Bayer *et al.*, 1998). Los celulosomas funcionan como estructuras exocelulares especializadas que catalizan la hidrólisis de celulosa y hemicelulosa (Hernández-Santoyo *et al.*, 1999). El principal componente del celulosoma es una gran subunidad de andamiaje, consistente en una proteína no catalítica la cual contiene por un lado módulos cohesina^P (generalmente múltiples copias) para la incorporación de las diferentes enzimas y por otro lado está sujeta a la proteína de anclaje a través de la cual se produce la unión a la pared celular. Existen además, unos módulos complementarios denominados dominios doquerina^Q o de acoplamiento que se unen fuertemente a los dominios cohesina dentro de cada subunidad del andamiaje. Esta interacción cohesina-doquerina media la integración de enzimas en el complejo y representa un carácter selectivo y estable dentro de este sistema. Por último, también suelen aparecer frecuentemente módulos de unión a carbohidratos (CBM) los cuales sirven para reconocer específicamente el sustrato polimérico y unirse a él (este tipo de dominios tienen especial importancia y se estudiarán en profundidad posteriormente (Apartado 1.4) (Bayer *et al.*, 1998).

^P Dominio cohesina: Responsable de la unión de las subunidades catalíticas dentro del celulosoma.

^Q Dominio doquerina: Ayuda a la integración de la enzima dentro del complejo celulosómico. Están unidos a las subunidades catalíticas por medio de péptidos de enlace.

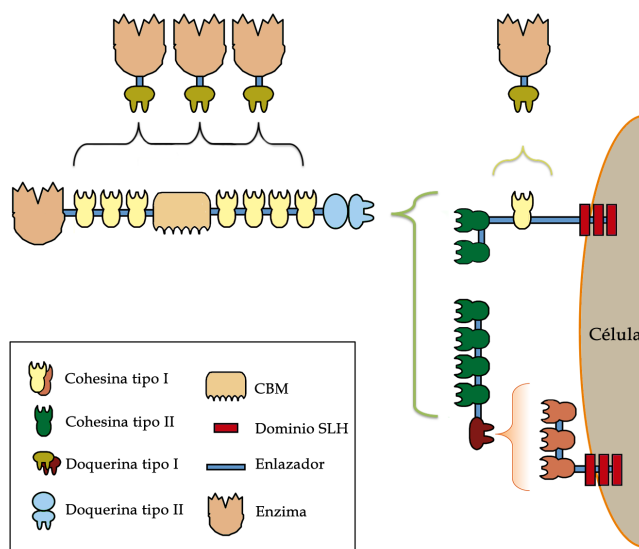


Figura 1.15. Representación esquemática de un celulosoma. Figura adaptada de (Hamberg *et al.*, 2014). Cortesía de Edward A. Bayer y Daniel B. Fried.

Aunque la mayoría de las enzimas que constituyen los celulosomas son celulasas, las xilanasas también se encuentran presentes dentro de los mismos (Pason *et al.*, 2006) así como otras GH. Todas estas enzimas, actuando de forma conjunta, hacen posible la degradación total de la lignocelulosa, gracias a las uniones específicas que establecen con su sustrato los diferentes tipos de CBM presentes tanto en la proteína de andamiaje como en las subunidades enzimáticas. Por analogía, el término xilanosoma ha sido propuesto para agregados de proteínas extracelulares en las cuales las xilanasas sean predominantes, aunque este no ha sido tan bien caracterizado como los celulosomas (Beg *et al.*, 2001; Jiang *et al.*, 2006).

1.3.1. Arquitectura molecular y clasificación

Teniendo en cuenta la arquitectura molecular, las xilanasas se pueden clasificar en dos tipos: xilanasas de dominio simple y multidominio. Las primeras, como su nombre indica, presentan únicamente un dominio catalítico mientras que las segundas poseen una estructura modular que tiene, además del dominio catalítico, uno o varios dominios accesorios unidos por segmentos conectores (Gilbert & Hazlewood, 1993; Gilkes *et al.*, 1991). En las enzimas modulares, el dominio catalítico define la especificidad y reactividad referente al sustrato, mientras que los dominios de unión dirigen la unión de la enzima al mismo. Estos dominios de afinidad tienen un plegamiento y una función independiente y pueden participar en la unión de celulosa (dominios de unión a celulosa, CBD) y/o xilano (dominios de unión a xilano, XBD). Para definirlos de un modo general se les denomina como dominios de unión de carbohidratos (módulos de unión a carbohidratos, CBM) (Boraston *et al.*, 2004).

Las GH están clasificadas en la actualidad en familias y estas, a su vez, en clanes o superfamilias^r. Este sistema de clasificación se desarrolló en 1991 en base a una comparación entre secuencias (Henrissat, 1991). Desde entonces, las 35 familias de GH identificadas inicialmente se han ampliado a 133 incluidas en la base de datos CAZy (*Carbohydrate-Active enZYmes*, <http://www.cazy.org/>) (Cantarel *et al.*, 2009; Lombard *et al.*, 2014) que está siendo continuamente actualizada (Lombard *et al.*, 2014). Su clasificación se basa en la comparación de secuencias pero también en características estructurales. De acuerdo con la base de datos CAZy, a la fecha de escritura de esta tesis, hay referenciadas 13 familias de GH que son consideradas endo-1,4- β -xilanasas (EC 3.2.1.8). Son las siguientes: 5, 8, 9, 10, 11, 12, 16, 26, 30, 43, 44, 51 y 62. En la Tabla 1.3, se indica el plegamiento, clan, mecanismo catalítico y los residuos implicados de las enzimas que presentan actividad xilanasas, mientras que en la Figura 1.16 se representan los plegamientos tipo de las familias de xilanasas más comunes:

Tabla 1.3. Familias de xilanasas con su plegamiento, clan, mecanismo catalítico y residuos implicados.

| Familia GH | Plegamiento | Clan | Mecanismo catalítico | Nucleófilo catalítico/base | Protón donador catalítico |
|------------|--|------|----------------------|----------------------------|---------------------------|
| 5 | Barril (β/α) ₈ | GH-A | Retención | Glutamato | Glutamato |
| 8 | (α/α) ₆ | GH-M | Inversión | Aspartato | Glutamato |
| 9 | (α/α) ₆ | GH-M | Inversión | Aspartato | Glutamato |
| 10 | (β/α) ₈ | GH-A | Retención | Glutamato | Glutamato |
| 11 | β -jelly roll | GH-C | Retención | Glutamato | Glutamato |
| 12 | β -jelly roll | GH-C | Retención | Glutamato | Glutamato |
| 16 | β -jelly roll | GH-B | Retención | Glutamato | Glutamato |
| 26 | (β/α) ₈ | GH-A | Retención | Glutamato | Glutamato |
| 30 | (β/α) ₈ | GH-A | Retención | Glutamato | Glutamato |
| 43 | 5-fold- β -propeller | GH-F | Inversión | Aspartato | Glutamato |
| 44 | (β/α) ₈ | GH-A | Retención | Glutamato | Glutamato |
| 51 | (β/α) ₈ | GH-A | Retención | Glutamato | Glutamato |
| 62 | Desconocido | GH-F | Desconocido | Desconocido | Desconocido |

(Código de colores. Naranja: Xilanasas clásicas. Azul: Xilanasas recientemente caracterizadas).

De todas estas enzimas, las más caracterizadas han sido las de las familias 10 y 11, las cuales no parece que estén especializadas en degradar un tipo concreto de xilano ya que son capaces de degradar tanto GXs como arabinoxilanos (Kolenova *et al.*, 2006; Pollet *et al.*, 2010). Estas dos familias de enzimas se consideran como las “xilanasas clásicas”. En general, actúan sobre el xilano poco ramificado y operan en conjunción con otras enzimas para la completa degradación del mismo. Recientemente se han descrito nuevas enzimas que requieren sustituciones específicas para su actividad. Por ejemplo, se ha visto que la familia GH30 requiere de decoraciones de metilglucurónico para la hidrólisis denominándose, por lo tanto, glucuronoxilanasas (Gallardo *et al.*, 2010; St John *et al.*, 2006). Además, recientemente se ha descrito la primera enzima específica por arabinoxilano (Correia *et al.*, 2011), la cual pertenece a la familia GH5.

^r Clan o superfamilia: Grupo de familias que se cree que comparten un antepasado común y presentan una misma estructura terciaria con conservación de los residuos catalíticos así como de su mecanismo.

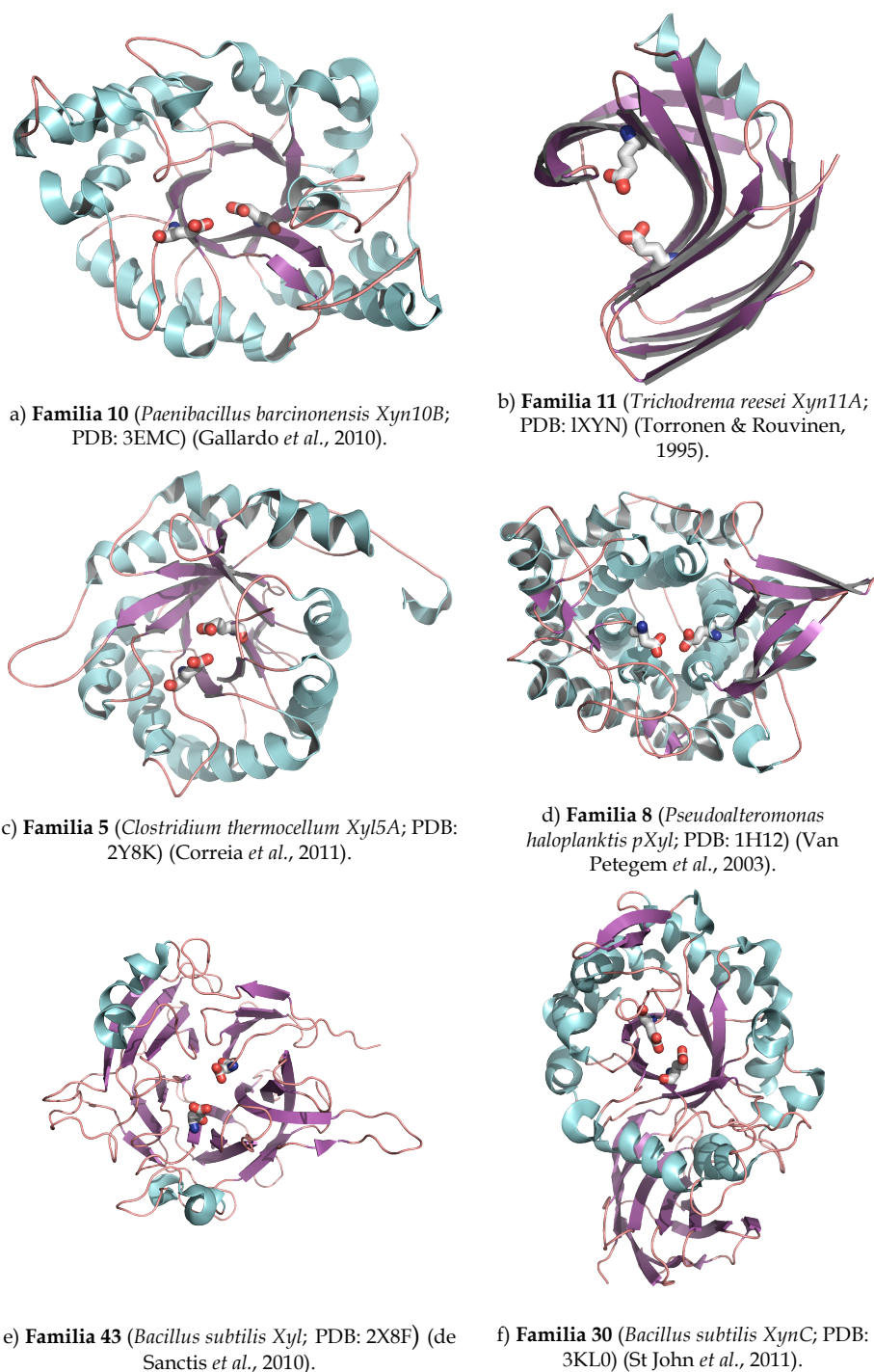


Figura 1.16. Representación de la estructura tipo de las siguientes xilanasas: a) GH10, (β/α)₈. b) GH11, β -jelly roll. c) GH5, (β/α)₈. d) GH8, (α/α)₆. e) GH43, 5-fold- β -propeller. f) GH30, (β/α)₈, con subdominio β_9 . (PDB: Banco de datos de proteínas (*Protein data bank*)).

Las enzimas estudiadas en este trabajo contienen dominios catalíticos de las familias GH10 y GH30. Por lo tanto, a continuación se describen las principales características de ambas familias.

1.3.2. Xilanasas GH10

Las xilanasas de la familia GH10 (o familia F) son principalmente endo-1,4- β -xilanasas (EC 3.2.1.8), pero en algunas ocasiones pueden ser también endo-1,3- β -xilanasas (EC 3.2.1.32) (Lombard *et al.*, 2014). Estas enzimas tienen un alto peso molecular, un bajo punto isoeléctrico (pI) y poseen una amplia especificidad de sustrato (Collins *et al.*, 2005). Pueden degradar tanto sustratos lineales como heteroxilanos. El análisis de los productos de hidrólisis tanto de GXs como de arabinoxilanos, muestra que las xilanasas de esta familia pueden atacar el enlace glicosílico cercano a una sustitución simple o doble de xilosa (Pollet *et al.*, 2010).

Las xilanasas GH10 son enzimas muy versátiles que liberan pequeños XOS e incluso xilosa a partir de sustratos de alta y baja DP (Van Campenhout *et al.*, 2007). Además de la característica xilanólítica, las xilanasas GH10 muestran actividad sobre sustratos derivados de glucosa como por ejemplo aril-celo-oligosacáridos, aunque no son activas en celulosa (Andrews *et al.*, 2000). Asimismo, pueden atacar enlaces β -1,4 que preceden a un enlace β -1,3 en xilanos mixtos, pero no los enlaces que siguen al β -1,3 hacia el extremo reductor (Biely *et al.*, 1997). Finalmente, también hidrolizan enlaces β -1,3 que estén acompañados a ambos lados por enlaces β -1,4 (Nerinckx *et al.*, 2004) y permiten el reemplazo de una o dos xilasas consecutivas por residuos de glucosa en el sustrato (Biely *et al.*, 1997).

Suelen presentar cuatro o cinco sitios de unión dentro del centro activo (Biely *et al.*, 1997). Son miembros del clan GH-A y su módulo catalítico presenta una estructura de barril (β/α)₈ también denominado *TIM-barrel* (que se refiere a la estructura de la enzima *Triose-phosphate isomerase*, Figura 1.17a). El plegamiento de las xilanasas GH10 está muy conservado y sólo se encuentran variaciones en los bucles que unen las hebras β interiores con las hélices que las rodean (Solomon *et al.*, 2007). Además, la forma global de este barril (β/α)₈ también puede variar desde una forma circular hasta una elíptica (Pollet *et al.*, 2010).

La topología del centro activo para estas xilanasas presenta un hendidura poco profunda en el extremo C-terminal de las láminas β del barril. Los dos glutamatos que participan en el mecanismo de retención se localizan cerca del extremo C-terminal en las hebras 4 y 7 (Figura 1.17) (Jenkins *et al.*, 1995).

Las estructuras cristalográficas de xilanasas GH10 y los complejos con sus sustratos han revelado que los subsitios -2, -1 y +1 están altamente conservados (Ducros *et al.*, 2000). De este modo, los aminoácidos en estos subsitios tienen un papel clave para el reconocimiento del sustrato y presentan numerosas interacciones entre la enzima y las unidades de xilosa. En la zona “glicón”^s, los

^s Glicón: Es el componente de azúcar del glicósido (se aloja en los sitios -n del centro activo). Aglicón: Es la porción de molécula variable que puede no pertenecer al grupo de los carbohidratos dentro del glicósido (se aloja en los sitios +n).

residuos de xilosa son reconocidos mediante enlaces de hidrógeno. Sin embargo, en la zona “aglicón”, el reconocimiento del sustrato está dominado por interacciones hidrofóbicas entre los anillos de xilopiranososa y residuos aromáticos (Figura 1.17) (Charnock *et al.*, 1998; Zolotnitsky *et al.*, 2004). La capacidad de las GH10 por hidrolizar y producir XOS pequeños está vinculada con la fuerza en el enlace de estos subsitios -2, -1 y +1 (Fujimoto *et al.*, 2004).

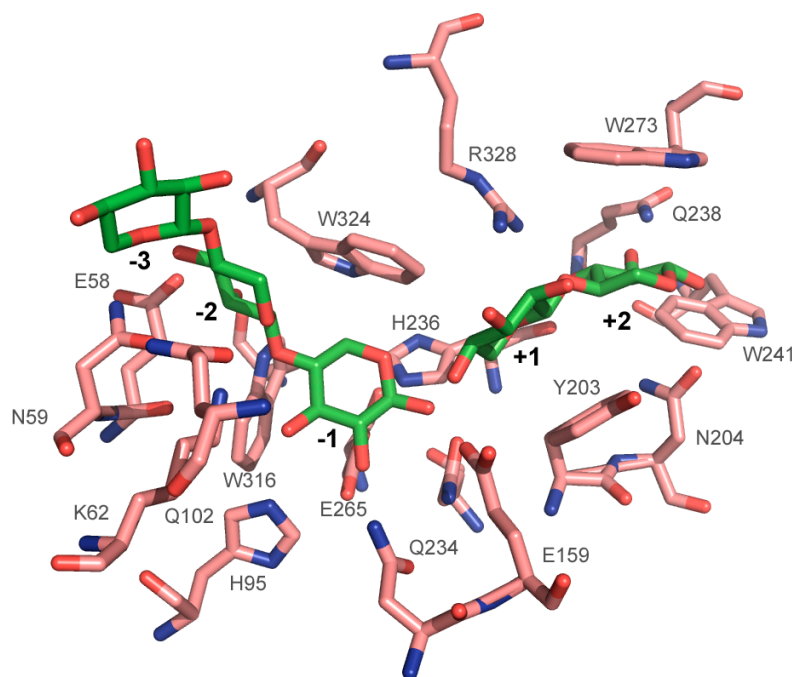


Figura 1.17. Esquema de interacciones de sitio de unión de la xilanasas de *Geobacillus stearothermophilus* en complejo con xilopentaosa (PDB: 1R87). Hay cinco residuos aromáticos próximos al centro activo que podrían estar formando parte de interacciones de *stacking* o “apilamiento” con los anillos del azúcar en diferentes subsitios: Trp316 en el subsitio -2, Trp324 en el -1, Tyr203 en el +1, Trp273 en el +2 y Trp241 en el +3. Estudios termodinámicos determinaron que para los subsitios -2 y -1, los residuos aromáticos tenían menor área de superficie de contacto (10-20 Å²) y además no eran paralelos, que es un requerimiento imprescindible para interacciones de *stacking*. Sin embargo, los residuos aromáticos en los subsitios +1 y +2 estaban orientados de forma paralela con el sustrato con una mayor superficie de área de contacto (>50 Å²), dando como resultado una interacción mucho más favorable (Zolotnitsky *et al.*, 2004).

1.3.3. Xilanasas GH30

La familia GH30 contiene enzimas con muy diferentes actividades, endo-β-1,4-xilanasas, β-glucosidasas, β-glucuronidasas, β-xilosidasas, β-fucosidasas, glucosilceramidasas, GAX endo-β-1,4-xilanasas o β-xilosidasas, entre otras (Lombard *et al.*, 2014).

Estas enzimas fueron inicialmente clasificadas dentro de la familia 5 de GH. Más recientemente se han descrito motivos estructurales comunes en varias xilanasas que han permitido catalogar una nueva familia bautizada como GH30

(St John *et al.*, 2010). Muchas xilanasas bacterianas de la familia GH30 están especializadas en la hidrólisis de xilanos que contienen residuos de GlcA o MeGlcA en la cadena lateral. A este tipo de enzimas se les denomina glucuronoxilanasas. Sin embargo, no todas las xilanasas GH30 muestran esta especificidad. Una excepción sería el caso de la xilanasas de hongo *Bispora* (Luo *et al.*, 2010). Aún así, existen xilanasas que requieren específicamente MeGlcA para su función, como son la xilanasas XynC de *Bacillus subtilis* y la xilanasas XynA de *Erwinia chrysanthemi* (Hurlbert & Preston, 2001; Nishitani & Nevins, 1991; St John *et al.*, 2006; Vrsanska *et al.*, 2007). Ambas enzimas presentan un único dominio y sus estructuras cristalinas han sido reportadas (St John *et al.*, 2011; Urbanikova *et al.*, 2011). Recientemente se han secuenciado dos enzimas de GH30 modulares; una de *Clostridium papirosolvens* (su caracterización bioquímica no se ha reportado hasta la fecha) y otra de *P. barcinonensis* (Valenzuela *et al.*, 2012), ambas con CBM adicionales. Esta última es una de las enzimas estudiadas en el trabajo objeto de esta Tesis, como luego se describirá.

El dominio catalítico de estas xilanasas tiene una estructura tipo barril (β/α)₈ perteneciente al clan A (común con las GH de la familia 5), unido a un subdominio único de estructura de láminas β de nueve hebras (Figura 1.16f). Las mayores diferencias en secuencia encontradas dentro de esta familia radican, precisamente, en esta estructura β (St John *et al.*, 2010). La estructura cristalina de la xilanasas XynC de *Bacillus subtilis* ha mostrado que este subdominio es capaz de unir el sustrato, por lo cual podría actuar como un dominio CBM “fusionado” al dominio catalítico (Figura 1.18). Sin embargo, como veremos más adelante, esta capacidad de unir el sustrato no es común a todos los miembros de la familia y está ausente en enzimas modulares como la estudiada en este trabajo.

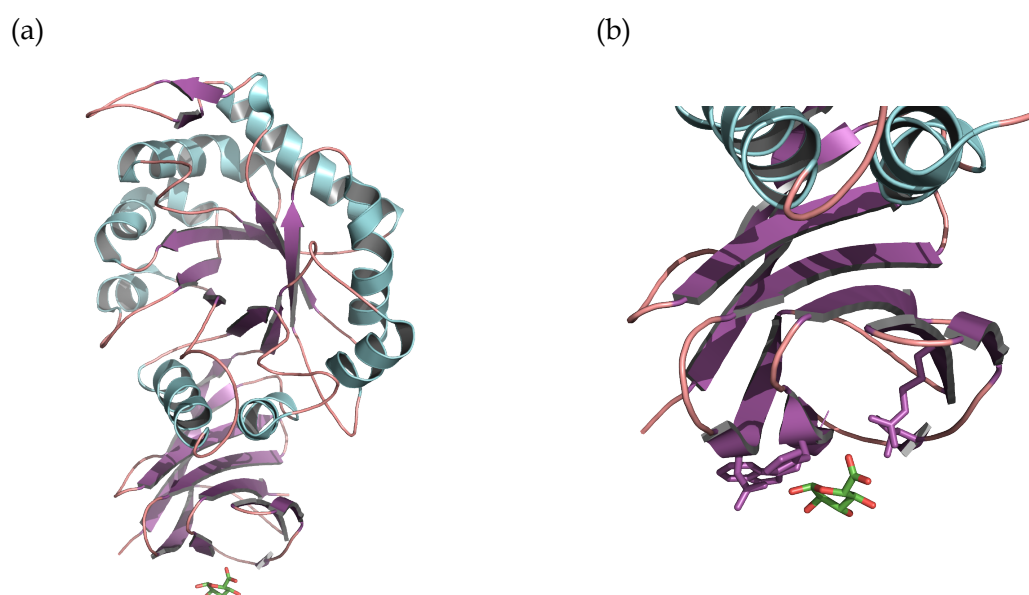


Figura 1.18. Imagen de la estructura cristalina de la glucuronoxilanasas GH30 de *Bacillus subtilis*, XynC en complejo con GlcA unido en el subdominio β de dicha enzima. (a) Representación de la molécula completa. (b) Representación de su sitio de unión ampliado (St John *et al.*, 2011).

Los estudios bioquímicos realizados con las xilanasas que se han caracterizado de esta familia nos han indicado que tienen actividad frente a GXs y no por arabinoxilanos (Valenzuela *et al.*, 2012). Además, la sustitución de MeGlcA es clave en la hidrólisis, por lo que esta está relacionada con el grado de sustitución. La reacción de ruptura se produce en el segundo enlace después de la ramificación, hacia el extremo reductor dentro de la cadena de xilano. De esta forma, se ha propuesto un mecanismo en el cual se requiere de la unión de un residuo de MeGlcA en el subsitio -2. Como resultado, los productos generados de la hidrólisis están formados por una molécula de MeGlcA unida al penúltimo residuo de xilosa a partir del extremo reductor. La resolución de la estructura tridimensional de la enzima GH30 de *Erwinia chrysanthemi*, junto con el análisis de su modo de acción (Vrsanska, 2007; Urbániková, 2011), permitió establecer por primera vez una hipótesis para explicar el mecanismo de reconocimiento del sustrato (Figura 1.19) en las xilanasas originalmente denominadas como “xilanasas apéndice-dependientes” (Nishitani & Nevins, 1991).

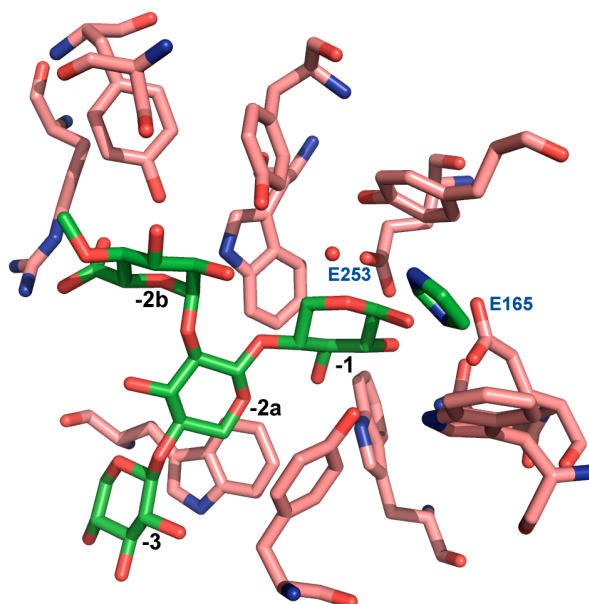


Figura 1.19. Esquema de interacciones de sitio de unión de la glucuronoxilanasas GH30 de *Erwinia chrysanthemi* en complejo con MeGlcA (PDB: 2Y24). Representación de los residuos que participan en la unión, destacando los aminoácidos catalíticos Glu165 y Glu253 (Urbanikova *et al.*, 2011).

1.4. Módulos de unión a carbohidratos

Como se ha adelantado previamente, muchas hidrolasas microbianas de la pared celular vegetal muestran una arquitectura modular en la cual el módulo catalítico está unido a uno o más módulos de unión a carbohidratos a través de segmentos conectores que le proporcionan una elevada flexibilidad (Gilbert, 2010).

Los CBM primeramente fueron definidos como módulos de unión a celulosa (CBD) ya que inicialmente se descubrieron varios módulos que unían este polisacárido (Gilkes *et al.*, 1988). Sin embargo, poco a poco se han ido descubriendo dominios adicionales de unión a azúcares que unen otros tipos de carbohidratos. Debido a esto, se propuso el término CBM el cual incluye a todos los módulos no-catalíticos de unión a azúcares derivados de GH.

CBD: Dominios no-catalíticos más abundantes en xilanasas. Han sido clasificados en varias familias (Tomme *et al.*, 1998). Participan en la unión de diferentes formas de celulosa amorfa y cristalina (Linder & Teeri, 1997).

XBD: Se caracterizó por primera vez un dominio XBD con muy baja afinidad por celulosa en comparación con su afinidad por xilano (Black *et al.*, 1995). Son dominios de unión que están presentes en muchas enzimas termófilas^t (Lee & Zeikus, 1993) y también se han encontrado en algunas xilanasas mesófilas^u (Blanco *et al.*, 1999). Se ha demostrado que promueven la unión a xilano (Sunna *et al.*, 2000).

CBM: El descubrimiento de módulos de unión a quitina^v y almidón hizo que se generalizara el término “módulo de unión a carbohidrato”, para todos los dominios que unen carbohidratos (Boraston *et al.*, 2004).

Aunque estén comúnmente asociados a GH, también se han identificado en liasas, oxidasas, glicosil transferasas y expansinas de la pared celular vegetal (Georgelis *et al.*, 2011). Aún así, en general los CBM están presentes en GH que degradan polisacáridos insolubles (Boraston *et al.*, 2004).

1.4.1. Clasificación

Del mismo modo que las GH, estos CBM están clasificados en la base de datos CAZy que, a fecha de escritura de este manuscrito, están divididos en 71 familias según su secuencia y especificidad de sustrato. Así, hay caracterizados módulos que son capaces de unir celulosa no cristalina, celulosa cristalina, quitina, β -1,3-glucanos y β -1,3-1,4-glucanos mixtos, xilano, manano, galactano y almidón. Además, hay otros que muestran tener especificidad tipo lectina y tienen afinidad por una variedad de glicanos en la superficie de la célula. De una forma general, se puede decir que existen CBM que son capaces de interactuar con casi todas las fuentes de carbohidratos existentes en la naturaleza (Tabla 1.4).

^t Organismo termófilo: Dícese del organismo que para su normal desarrollo requiere temperaturas elevadas.

^u Organismo mesófilo: Se aplica a organismos que se desarrollan a temperatura y humedad medias.

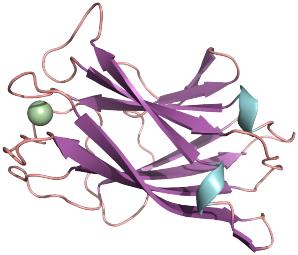
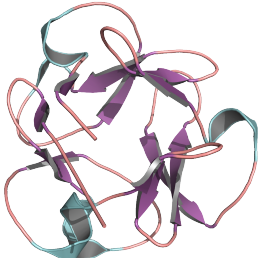
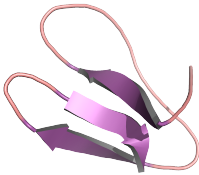
^v Quitina: n. f. Bioquím. Es el principal polisacárido de las paredes celulares de hongos y del exoesqueleto de los artrópodos.

Tabla 1.4. Carbohidratos que interaccionan con las diferentes familias de CBM (CAZy)^w.


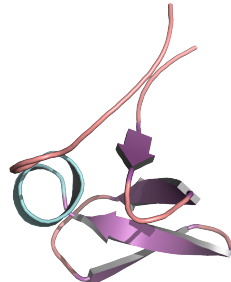
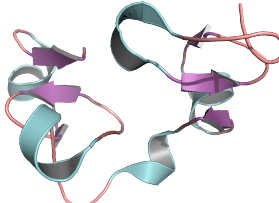
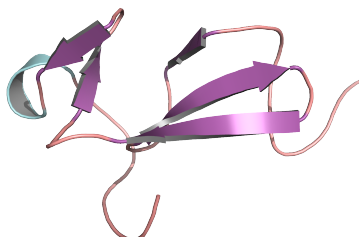
| Carbohidrato | CBM |
|---|--|
| Celulosa | 1, 2, 3, 4 (amorfa), 6 (amorfa), 8, 9, 10, 16, 17 (amorfa), 28 (amorfa), 30, 37 (cristalina), 44, 46, 49 (cristalina), 63, 64. |
| Xilano | 2, 4, 6, 9, 13, 15, 22, 31, 35, 36, 37, 44, 54, 59, 60. |
| Pared celular vegetal (β -glucanos, glucanos mixtos, pectinas, mananos, etc) | 4, 6, 11, 13, 16, 22, 23, 27, 28, 29, 32, 35, 39, 42, 43, 52, 56, 59, 61, 62, 65, 67. |
| Quitina | 1, 2, 3, 5, 12, 14, 18, 19, 37, 50, 54, 55. |
| α -glucanos (almidón, glucógeno, etc) | 20, 21, 24, 25, 26, 34, 41, 45, 48, 53, 58, 68, 69. |
| Glicanos de mamíferos | 32, 40, 47, 51, 57. |
| Otros | Pared celular de bacterias: 35, 39, 50. Pared celular de levaduras: 54. Fructanos: 38, 66. |

En 2004, Boraston y colaboradores (Boraston *et al.*, 2004) clasificaron los CBM en 7 familias estructurales según el plegamiento que presentaban. Ejemplos de CBM pertenecientes a cada tipo de plegamiento se representan en Tabla 1.5:

Tabla 1.5. Representación de los plegamientos para cada familia de CBM.

| Plegamiento | Ejemplo estructura proteína | Familia CBM | Código PDB |
|---------------------------------|---|-------------|--|
| Sándwich β (Familia 1) |  | CBM3 | 1NBC (Tormo <i>et al.</i> , 1996) |
| Trébol β (Familia 2) |  | CBM13 | 1MC9 (Notenboom <i>et al.</i> , 2002) |
| Nudo de cisteína (Familia 3) |  | CBM1 | 1CBH (Kraulis <i>et al.</i> , 1989) |

^w CBM7, se ha borrado de la base de datos. CBM33, ahora está dentro de la familia AA10 (actividades auxiliares), es una polisacárido monooxigenasa, dependiente de cobre.

| | | | |
|--|--|-------|---|
| β-Única (Familia 4) |  | CBM5 | 1AIW (Brun <i>et al.</i> , 1997) |
| Plegamiento OB (Familia 5) |  | CBM10 | 1E8R (Raghothama <i>et al.</i> , 2000) |
| Plegamiento tipo heveína (Familia 6) |  | CBM18 | 1EN2 (Saul <i>et al.</i> , 2000) |
| Única (contiene plegamiento tipo heveína) (Familia 7) |  | CBM14 | 1DQC (Suetake <i>et al.</i> , 2000) |

Además, para proporcionar información funcional adicional, los CBM se clasificaron fundamentalmente en tres tipos –A, B, y C– según unieran superficies cristalinas de celulosa, cadenas de glicanos u oligosacáridos cortos, respectivamente (Figura 1.20) (Boraston *et al.*, 2004; Hashimoto, 2006).

Desde entonces, se han obtenido una gran cantidad de datos estructurales y funcionales fundamentalmente para miembros de los tipos B y C, lo cual ha aportado nuevos detalles que permiten conocer mucho mejor esta familia de proteínas. En base a ello, Gilbert propuso en 2013 una nueva clasificación más completa (Gilbert *et al.*, 2013):

- Tipo A: El sitio de unión comprende una plataforma plana hidrófoba rica en residuos aromáticos los cuales establecen la unión con sustratos planos policristalinos de celulosa o quitina (Figura 1.20a). Este tipo de CBM se encuentra también en expansinas, proteínas que rompen la interfase celulosa-hemicelulosa. Un ejemplo es la expansina de *Bacillus subtilis* (CBM63; PDB: 4FER) en complejo con celohexaosa (Georgelis *et al.*, 2012). Este tipo de CBM (Jervis *et al.*, 1997) es muy diferente a los otros dos grupos.
- Tipo B: Une cadenas internas (tipo “endo”) y, hasta la fecha, son los más abundantes. El sitio de unión es un canal alargado con una hendidura que

puede alojar cadenas de cuatro o más unidades de monosacáridos (Figura 1.20b).

- **Tipo C:** Tiene la capacidad de unir extremos terminales (reductores o no-reductores, unión tipo “exo”) (Gilbert *et al.*, 2013). El sitio de unión es un bolsillo pequeño que puede alojar oligosacáridos cortos, generalmente de uno, dos o tres monosacáridos (Figura 1.20c). Dentro de este tipo se encuentran algunas lectinas.

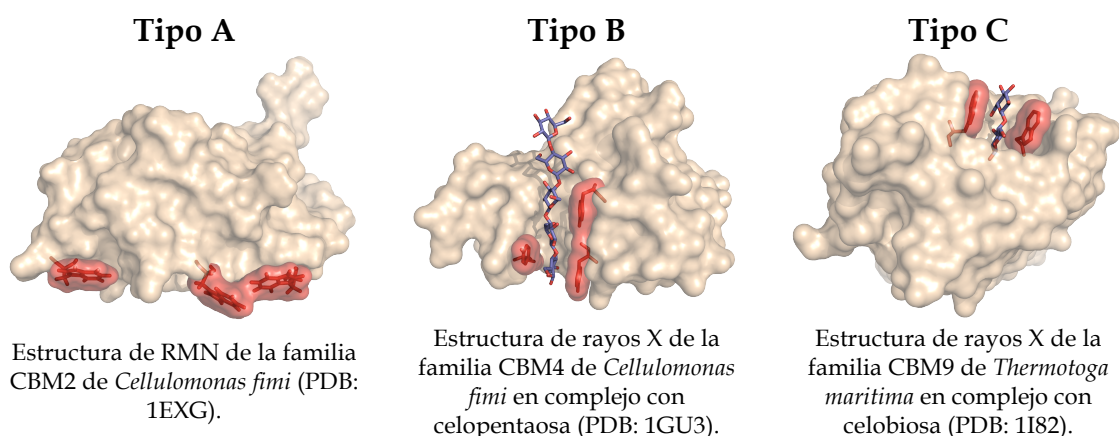


Figura 1.20. Representación de cada uno de los tipos de CBM, A, B y C, con sus sitios de unión resaltados mediante una superficie en color rojo. (RMN: resonancia magnética nuclear).

La Figura 1.21 esquematiza el modo de unión al sustrato de los distintos CBM.

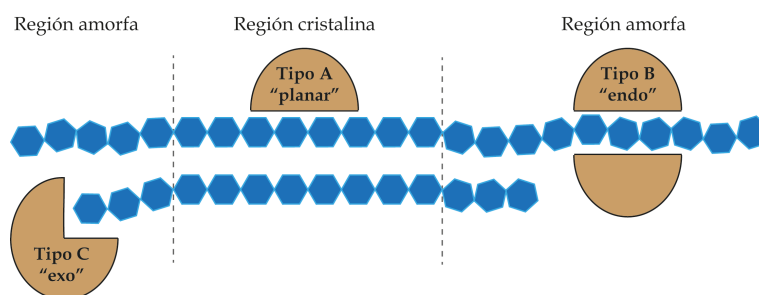


Figura 1.21. Diferentes modos de unión a sustrato de CBM tipo A, B y C.

Una característica importante a destacar es que algunas familias de CBM son capaces de reconocer y acomodar cadenas sustituidas, como puede ser, por ejemplo, el XG. Así, algunos datos estructurales han proporcionado información sobre el mecanismo de especificidad de los CBM tipo B, indicando que contienen una plataforma hidrófoba extendida que es capaz de establecer contactos apolares significativos con las cadenas laterales de xilosa, mientras que también puede interactuar con las moléculas de glucosa de la cadena principal del sustrato. En la Figura 1.22 se puede apreciar la topología característica de este tipo de unión (Luis *et al.*, 2013). De hecho, dicho mecanismo de unión proporciona flexibilidad en el reconocimiento de XG donde el registro estándar de XXXG (siendo X: xilosa y G: glucosa) puede variar

entre las diferentes formas del polisacárido (Fry *et al.*, 1993). De esta forma, se demostró que el reconocimiento de la cadena lateral es tanto o mas importante que los contactos polares que pueda haber con el resto de la cadena principal (Luis *et al.*, 2013).

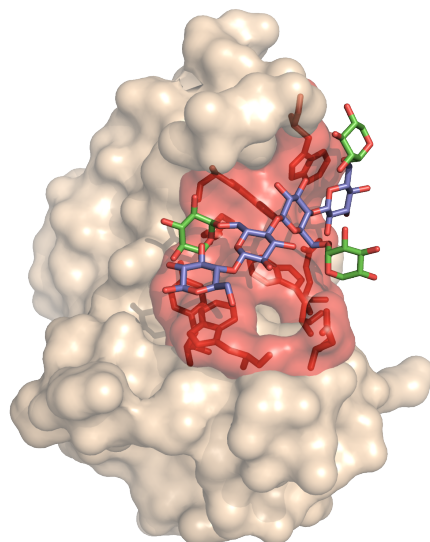


Figura 1.22. CBM de la familia 65 de *Eubacterium cellulosolvens* (PDB: 2YPJ) en complejo con un fragmento de XG.

1.4.2. Función

Los CBM no suelen sufrir cambios conformacionales cuando se unen al ligando. Sin embargo, las enzimas multimodulares en su conjunto suelen ser muy flexibles y pueden presentar variaciones al unirse a su sustrato. De este modo, el módulo adicional puede dirigir y orientar al catalítico a zonas concretas del polímero (Ficko-Blean *et al.*, 2009).

Habitualmente, la afinidad de un CBM por su sustrato diana en la planta es muy baja (su constante de unión, K , es aproximadamente $100 \mu\text{M}$). Sin embargo, algunas enzimas contienen múltiples copias de CBM que muestran la misma especificidad y, en esas proteínas, los efectos concertados entre dichos módulos lleva a un aumento de la afinidad por polisacáridos. Esto es muy interesante ya que demuestra que la naturaleza ha preferido desarrollar la duplicación de CBM para incrementar la afinidad antes que incrementar las interacciones del ligando dentro de un solo CBM (Gilbert, 2010). Estos módulos presentan cuatro papeles funcionales principalmente:

- a) Efecto de proximidad: Incrementan la concentración de la enzima en la superficie del sustrato. Esto se traduce en una degradación del mismo más rápida y eficiente (Boraston *et al.*, 2004).

- b) Efecto de orientación: Los CBM pueden dirigir a la enzima a distintas partes del sustrato (extremo terminal reductor, terminal no reductor, cadenas intermedias de polisacárido, etc.), dependiendo de la topología del sitio de unión (Herve *et al.*, 2010).
- c) Efecto perturbador: Se ha demostrado que algunos CBM son capaces de alterar la superficie de polisacáridos altamente empaquetados (por ejemplo fibras de celulosa) facilitando, de este modo, que esté más expuesto al módulo catalítico y así se logre una degradación mas eficiente (Din *et al.*, 1991; Gourlay *et al.*, 2012).
- d) Adherencia: Existen CBM que se adhieren a componentes de superficie de la pared celular bacteriana y realizan su actividad catalítica sobre un carbohidrato vecino externo. Por ejemplo, un CBM de la familia 35 puede interaccionar con azúcares que tienen GlcA en la pared celular de *Amycolatopsis orientalis* mientras que el módulo catalítico es activo en el quitosano^x externo. Este quitosano puede proceder por ejemplo de la pared celular de especies competidoras de hongos del suelo (Montanier *et al.*, 2009).

Estudios recientes han mostrado que grupos de CBM que aparentemente exhiben las mismas especificidades frente a polisacáridos aislados, pueden mostrar un comportamiento diferente frente a los polímeros equivalentes presentes en paredes celulares intactas (Herve *et al.*, 2010). Esto se debe tener en cuenta cuando se realizan estudios de afinidad sobre sustratos sintetizados en el laboratorio (ver Apartado 1.4.6).

En general, la correlación entre el ligando que es capaz de reconocer un CBM y el sustrato específico que actúa sobre su dominio catalítico es muy estrecha. Sin embargo, existen algunas excepciones a esta regla. Por ejemplo, los CBM de unión a celulosa, que siendo dominios de unión a este polisacárido, forman parte frecuentemente de enzimas que hidrolizan xilanos, mananos y pectinas (Herve *et al.*, 2010). Cada vez se identifican más enzimas donde existe una disparidad de especificidad entre el módulo catalítico y su CBM, lo cual indica que queda aún mucho por conocer acerca del papel fisiológico que juegan estas proteínas. Las relaciones filogenéticas de las secuencias de los CBM pueden ayudarnos a comprender la función en estos casos. Sin embargo, debido a la escasez de datos procedentes de CBM incluidos en la proteína completa, es difícil predecir en muchos casos su papel preciso. Actualmente, se ha propuesto que la química de los segmentos conectores o *linkers* así como las plataformas hidrofóbicas superficiales son factores muy importantes que hay que tener en cuenta para entender el papel exacto de los CBM en la regulación y modulación de la actividad enzimática (Abbott & van Bueren, 2014).

^x Quitosano: También llamado chitosán, es un polisacárido lineal compuesto de cadenas distribuidas aleatoriamente de β -(1,4)-D-glucosamina y N-acetil-D-glucosamina. Esta sustancia se descubrió en el año 1859. Se puede usar en agricultura como fungicida y en la industria vitivinícola para evitar el deterioro del vino. En medicina se usa a veces como añadido en vendajes para reducir el sangrado y disminuir la cantidad de infecciones.

1.4.3. Papel del hidrógeno y del calcio

Aunque la orientación y posición de los residuos aromáticos en el sitio activo es clave para una primera orientación del sustrato, otras interacciones, por ejemplo enlaces de hidrógeno y/o coordinación con el calcio, tienen un papel fundamental en el reconocimiento del ligando.

Los carbohidratos son moléculas anfipáticas las cuales, debido a la gran cantidad de grupos hidroxilo presentes en su estructura, tienen una gran capacidad para formar puentes de hidrógeno con los residuos polares de los sitios de unión dentro de las proteínas. Dependiendo del tipo de CBM (A, B o C), la interacción del sustrato es diferente. Así, la mutación de los residuos polares por alanina para los tipos B y C tienen una disminución de la afinidad mucho menor que para el tipo A. Además, debe tenerse en cuenta que esta disminución de afinidad puede deberse no sólo a la pérdida de los enlaces de hidrógeno ya que también puede influir el pequeño cambio estructural que sufren los sitios de unión (Boraston *et al.*, 2004).

El calcio tiene un papel clave en la interacción de muchos CBM con sus ligandos diana ya sea por el mantenimiento de la conformación del sitio de unión (calcio estructural) o por la coordinación directa con el propio carbohidrato (calcio funcional). Así, el ión calcio puede participar directamente en la unión de los carbohidratos, como se ha visto en las familias 35 y 36 (Bolam *et al.*, 2004; Jamal-Talabani *et al.*, 2004) coordinándose, por ejemplo en el caso de la familia 36, con los O2 y O3 de la cadena de xilano. Hasta 2009 no se conocieron los mecanismos de reconocimiento calcio-carbohidrato en la familia CBM35, los cuales se describen en el apartado siguiente (Montanier *et al.*, 2009). Normalmente, para los CBM que reconocen mono- o disacáridos el calcio suele presentar una unión directamente al ligando (Correia *et al.*, 2010).

Finalmente, el calcio también puede estar involucrado en la oligomerización en ciertos CBM como es el caso del CBM62 de *Clostridium thermocellum*, que forma un dímero dependiente de calcio. Esto tiene especial importancia biológica, si se tiene en cuenta la proteína de la que forma parte, CtXyn5A, que es un componente del celulosoma. Así, la dimerización de CtCBM62 podría reunir dos unidades de CtXyl5A en la misma molécula del celulosoma (Montanier *et al.*, 2011).

A continuación se detallan las principales características de las familias CBM9, CBM22 y CBM35, familias a las que pertenecen los dominios no catalíticos presentes en las enzimas objeto de este estudio.

1.4.4. Familias CBM22 y CBM9

Los primeros dominios identificados dentro de lo que hoy se denomina familia CBM22 estaban incluidos en enzimas procedentes de organismos

termófilos. Fueron inicialmente considerados como “dominios termoestables” (X6) (Charnock *et al.*, 2000) gracias a la aparente capacidad que tenían de proporcionar estabilidad al dominio catalítico. Sin embargo, todas las enzimas reportadas correspondían a xilanasas con dominios catalíticos GH10 y GH11 (Sunna *et al.*, 2000), lo cuál sugería una funcionalidad común. Además, se identificaron dominios homólogos en enzimas procedentes de bacterias mesófilas (Blanco *et al.*, 1999; Clarke *et al.*, 1996). Todo ello hizo que se generalizara su función como módulos de unión a carbohidratos y se incluyeron en la familia CBM22. Se encontró que estos dominios tenían afinidad tanto por polisacáridos solubles como insolubles, aunque parecen potenciar de manera especial la actividad catalítica sobre sustratos insolubles (Araki *et al.*, 2004; Charnock *et al.*, 2000).

La familia **CBM22** tiene como característica principal la unión a xilanos, si bien algunos han demostrado también una alta afinidad por β -glucanos o incluso por celulosa cristalina (Khan *et al.*, 2013). La primera estructura resulta por cristalografía de rayos X de un CBM22 fue la del dominio CBM22-2 de la Xyn10B de *Clostridium thermocellum* (Cth) (Charnock *et al.*, 2000). El análisis de la misma ha desvelado que tres residuos aromáticos y dos polares (Trp53, Tyr103, Tyr136, Arg25 y Glu138) son claves para mantener la integridad estructural del sitio de unión (Xie *et al.*, 2001). CthXyn10B es una enzima bifuncional con actividad xilanasas/esterasa y una estructura de dominios CBM22-1-GH10-CBM22-2-DOCK-CE1. Recientemente se ha resuelto la estructura cristalina de la parte N-terminal que contiene los dominios CBM22-1-GH10 (Najmudin *et al.*, 2010). El módulo CBM22-1 presenta una topología muy similar al CBM22-2 previamente resuelto (Figura 1.23), aunque, sin embargo, no presenta los residuos aromáticos incluidos en el motivo de unión a xilano.

El análisis de las secuencias de los CBM22 mostró que menos de la mitad de los módulos presentan los cinco motivos de unión, por lo que algunos deben presentar funciones desconocidas hasta el momento, uniendo sustratos aún por identificar. Otra hipótesis es que unan zonas de la interfase entre los distintos componentes de la pared, siendo estas interacciones difíciles de detectar mediante el análisis de unión a polisacáridos aislados. Por ejemplo, ciertas proteínas presentes en plantas llamadas expansinas, tienen como diana la interfase entre las microfibrillas de celulosa con hemicelulosas de la pared (Xie *et al.*, 2001). Por tanto, es necesario obtener más información para entender en profundidad los mecanismos de reconocimiento de la familia CBM22.

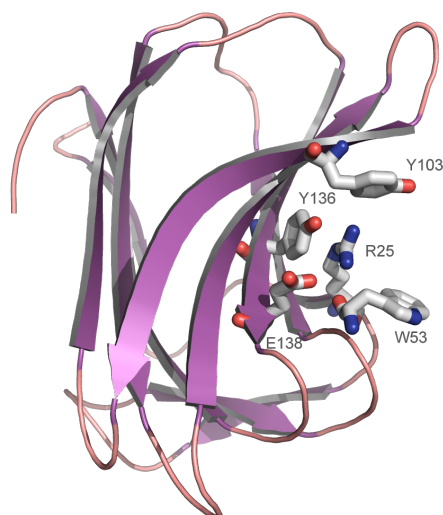


Figura 1.23. Detalle de los residuos claves en el sitio de unión en el caso del dominio CBM22-2 de *CthXynB* (PDB: 1DYO).

La mayor parte de módulos identificados de la familia CBM22 se encuentran en enzimas modulares que contienen dominios catalíticos GH10. En muchos casos, las enzimas contienen además módulos de la familia **CBM9**, que son proteínas de aproximadamente 170 residuos. Hasta ahora, únicamente se han encontrado formando parte de xilanasas. El único CBM9 resuelto por cristalografía de rayos X es el perteneciente a la xilanasa XynA de *Thermotoga marítima* (Notenboom *et al.*, 2001). Esta enzima contiene un tándem de dos CBM9 en el extremo C-terminal (CBM9-1-CBM9-2). Notenboom y colaboradores propusieron desglosar la familia CBM9 en dos subfamilias: 9a y 9b (CBM9-1 y CBM9-2, respectivamente). En los estudios de afinidad de ambas subfamilias se ha concluido que mientras CBM9a no tiene afinidad por carbohidratos (no presenta el motivo de unión), CBM9b puede unir celulosa cristalina y amorfa, xilano y un amplio rango de azúcares pequeños como mono- y disacáridos y celo- y XOS. Esto plantea la cuestión de cual puede ser el papel del CBM9a. Después de descartar la termoestabilidad como algo exclusivo para un dominio de estas características, se propuso una función de “espaciador” entre el dominio catalítico y el CBM9b, es decir, puede desempeñar un papel de *linker* (Boraston *et al.*, 2001; Notenboom *et al.*, 2001). El sitio de unión de esta familia CBM9 se encuentra en una hendidura profunda bloqueada en un extremo. De este modo, es capaz de unir específicamente el extremo reductor de los azúcares (Notenboom *et al.*, 2001).

1.4.5. Familia CBM35

Esta familia contiene módulos de aproximadamente 130 residuos. Suelen encontrarse junto a GH de las familias GH26, mananasas de la GH5 (Bolam *et al.*, 2004; Sunna, 2010; Taylor *et al.*, 2005) xilanasas de la GH30 (Valenzuela *et al.*, 2012), presentando afinidad por galactomanano, glucomanano, manano o GX (Ghosh *et al.*, 2013). Generalmente, CBM35 acomoda la cadena de polisacárido

mediante interacción hidrofóbica con residuos aromáticos, adoptando una configuración de tipo C.

Como se discutió anteriormente, los CBM generalmente presentan la misma especificidad de sustrato que los módulos catalíticos a los que acompañan. Una excepción importante son las xilanasas que presentan dominios de unión a celulosa. Recientemente, se ha reportado que ciertos dominios de la familia CBM35 presentan especificidad común e independiente de la actividad enzimática en la que están involucrados. En 2009, Montanier y colaboradores observaron que un grupo de CBM35, con una identidad de secuencia mayor del 38 %, pertenecían a enzimas cuyos módulos catalíticos tenían afinidad por sustratos muy diferentes. Aunque la mayor parte de los dominios forman parte de hidrolasas, uno de ellos presenta una función de adhesión celular (Cote *et al.*, 2006). Aún así, estos CBM35 tienen la capacidad común de unir ligandos muy similares. Por ejemplo, todos unen ácido $\Delta 4,5$ -anhidrogalaacturónico ($\Delta 4,5$ -GalA) y dos de ellos, además, GlcA, siendo ambos productos de degradación de la pectina. Por lo tanto, tienen la capacidad de unir productos de reacción generados por pectato liasas^v. Así, estarían dirigiendo a la enzima a zonas de la pared celular vegetal que está en proceso de degradación y/o es susceptible de ella (Montanier *et al.*, 2009).

Hasta la fecha, únicamente hay resuelta la estructura cristalográfica de un CBM35 perteneciente a una xilanasas. Se trata de Xyn-CBM35 de la xilanasas Xyn10B de *Cellvibrio japonicus* (Montanier *et al.*, 2009). Esta Xyn-CBM35 es capaz de unir xilano, pero no interacciona con XOS sin sustituir (Boraston *et al.*, 2004). Por lo tanto, su capacidad para unir este polímero es probable que dependa de la presencia de ramificaciones de GlcA en la cadena de xilosas, de tal forma que se necesite de estas “decoraciones laterales” en el xilano para que estos dominios se puedan unir al mismo. Sin embargo, se ha comprobado que la proteína no reconoce MeGlcA. La proporción de GlcA y MeGlcA en xilanos depende del equilibrio entre la velocidad de síntesis de GX y la velocidad de metilación de las cadenas laterales de este (Pena *et al.*, 2007). Aunque la razón biológica de dirección hacia el ácido urónico sin metilar no está clara, Montanier especuló que las regiones que contienen GlcA en vez de MeGlcA pueden tener una estructura más abierta, de tal forma que el dominio CBM35 dirige la enzima a zonas que son más accesibles y, por tanto, más susceptibles de degradación por el dominio catalítico (Montanier *et al.*, 2009).

^v Pectato liasa: Enzima que rompe los enlaces glicosídicos α -(1,4) del ácido poligalaacturónico, el componente principal de la pectina y de la lámina media de la pared celular vegetal.

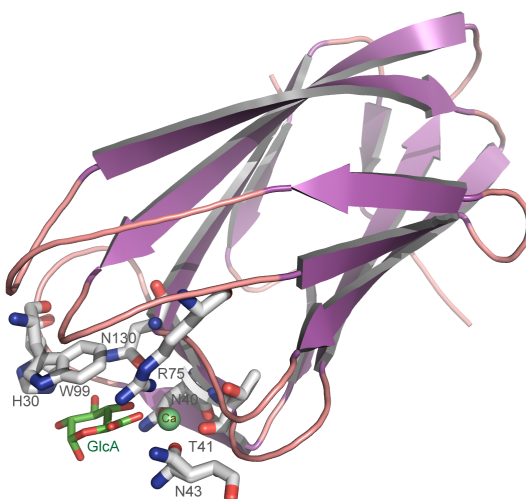


Figura 1.25. Detalle de los residuos claves en el sitio de unión en el caso del dominio CBM35 de CjXyn10B (PDB: 2W87).

1.4.6. Ensayos de unión *in vivo* mediante inmunohistoquímica

La mayoría de los ensayos de unión de los CBM a posibles sustratos han sido realizados *in vitro* utilizando polímeros aislados de su entorno natural. Por este motivo, es difícil realizar estudios estructura-función que conduzcan a conclusiones precisas de los mecanismos moleculares que determinan su especificidad. Recientemente se han realizado estudios de unión *in vivo* que ilustran muy bien, por una parte la gran heterogeneidad de los posibles sustratos naturales de cada CBM y, por otra, la enorme dependencia de los resultados obtenidos en función del entorno en que se analiza un determinado sustrato polimérico dentro de la pared celular vegetal altamente heterogénea. Así, mediante microscopía de inmunofluorescencia, McCartney y colaboradores analizaron la especificidad por sus sustratos de un grupo de seis familias de CBM utilizando distintos tipos de paredes celulares, mostrando considerables diferencias de especificidad y afinidad de unión en función del distinto tipo de material utilizado (McCartney *et al.*, 2006). Los investigadores que desarrollaron estos análisis tomaron como referencia el anticuerpo monoclonal LM11, el cual se sabe que une xilanos no sustituidos y arabinoxilano dentro de la pared secundaria de dicotiledóneas, pero no muestra afinidad dentro de la pared primaria de las mismas (McCartney *et al.*, 2005).

Este trabajo demuestra la importancia de investigar el papel de los CBM *in vivo* para determinar la diferente especificidad de los mismos en la pared primaria o secundaria. De este modo, pueden ser utilizados como sonda para investigar la estructura molecular de las plantas y proporcionar información de los polisacáridos que la forman en el contexto general de pared celular vegetal.

1.5. Aplicaciones biotecnológicas de las xilanasas

Las xilanasas y los microorganismos que las producen se usan cada día más en la gestión de desechos para degradar el xilano hasta productos químicos que tienen múltiples aplicaciones (Motta *et al.*, 2013). Un problema importante que presenta la degradación del xilano es su gran complejidad estructural. Por ello se requiere de una amplia variedad de xilanasas, capaces de reconocer y actuar frente a sustratos muy diversos (Collins *et al.*, 2005). Una de las aplicaciones biotecnológicas más importantes es su uso como biocatalizadores, reduciendo de este modo el impacto ambiental dentro de la industria maderera y de agroalimentación. Más recientemente se ha incrementado su uso en biomedicina.

1.5.1. Biomedicina e industria farmacéutica

Los términos “microflora” o “microbiota” intestinal hacen referencia a la flora microbiana que habita el tracto gastrointestinal. Los microorganismos presentes en el intestino desempeñan un papel fundamental en el correcto funcionamiento del organismo, no sólo desde el punto de vista de la digestión, sino también son claves dentro del sistema inmunitario. La regulación de la composición de esta microbiota intestinal es un tema que despierta gran interés en la actualidad para el desarrollo de una medicina preventiva personalizada (Clemente *et al.*, 2012) ya que se ha relacionado la composición de la microbiota con una mayor predisposición a sufrir ciertas enfermedades (Rodríguez *et al.*, 2015). Los prebióticos son componentes alimenticios no digeribles que estimulan el crecimiento de las bacterias intestinales de manera selectiva. De este modo, pueden dirigirse específicamente a ciertas poblaciones de bacterias con resultados muy beneficiosos para la salud. Un tipo de prebióticos muy utilizados en la actualidad son los XOS (Carvalho *et al.*, 2013) y en concreto, XOS procedentes del arabinoxilano (AXOS). Estos productos se obtienen por la acción de xilanasas. Grootaert y colaboradores (Grootaert *et al.*, 2009), realizaron un estudio en el cual concluyeron que en comparación con inulina, AXOS tenía un mayor potencial para trasladar una parte de la fermentación de los azúcares hacia la parte distal del colon. Por lo tanto, sugirieron este producto como un buen candidato para reducir los niveles de colesterol debido a su influencia beneficiosa en el metabolismo de la grasa. Además propusieron que, debido a las diferentes características prebióticas de AXOS en comparación con la inulina, sería interesante probar un producto combinado que contuviera tanto AXOS como inulina, el cual podría mostrar una mayor eficacia en un funcionamiento saludable del colon.

Otra aplicación de estos XOS es que son útiles para prevenir las caries dentales (Verma & Satyanarayana, 2012). Además, la xilosa que resulta de la degradación del xilano se puede transformar en xilitol, un edulcorante muy empleado en la industria farmacéutica (Motta *et al.*, 2013).

1.5.2. Conversión de la biomasa

De acuerdo con Taherzadeh (Taherzadeh & Karimi, 2008), el etanol es el combustible más importante en términos de volumen y de valor de mercado. Además, debido a la crisis del petróleo fósil, ha sido identificado como una de las alternativas más potentes. Dentro de este contexto, las xilanasas son enzimas claves en la degradación del material de deshecho lignocelulósico (junto con otras enzimas como celulasas, etc.), el cual se emplea como materia prima en este proceso. El primer paso sería la deslignificación de la lignocelulosa, para liberar celulosa y hemicelulosa. El segundo, la despolimerización de cada uno de estos sustratos para producir los azúcares libres, seguido de una fermentación para llegar a la producción final de etanol. Durante el proceso, además, se generan XOS, los cuales son útiles como prebióticos, ya mencionados en el apartado anterior.

1.5.3. Otros usos industriales

a) Bioblanqueamiento del papel

En el proceso de fabricación del papel, para lograr una pulpa de calidad se requiere separar de forma efectiva las fibras de celulosa de la lignina sin modificar las propiedades de la misma. Una etapa clave es el blanqueamiento de la pulpa. Clásicamente, se usaban compuestos que generaban residuos nocivos y contaminantes. Las enzimas pueden desbancar a estas sustancias y, debido a que son más específicas y biodegradables, no causan perjuicio al medioambiente. Viikari fue quien atribuyó a las xilanasas un papel en la industria del blanqueamiento del papel, reduciendo así el consumo de cloro un 25-30 % (Viikari *et al.*, 1994). Existe una demanda creciente de xilanasas termoestables ya que se requiere de esta propiedad para la producción de pulpa, la cual se realiza en condiciones de alta temperatura y pH alcalinos (Verma & Satyanarayana, 2012).

b) Industria alimentaria

En la manufactura del pan, las xilanasas contribuyen en la solubilización de la fracción de arabinoxilano presente en la masa, aumentando su volumen y calidad (de Vries & Visser, 2001). También son de gran utilidad en el procesado de cereales. El tratamiento de la harina de trigo con estas hace que se reduzca la absorción de agua lo cual genera grandes cambios en su textura (Verma & Satyanarayana, 2012). Una aplicación conjunta con pectinasas, celulasas y amilasas, es que pueden emplearse en la clarificación de zumo, cerveza y vino. Además, también son útiles en la extracción de aceites vegetales, café y almidón (Motta *et al.*, 2013).

c) Industria de piensos

Los XOS mencionados anteriormente también se usan en la industria agrícola como piensos para animales. El uso de xilanasas en las dietas de animales no rumiantes ha sido la solución para que aumentara su valor nutricional. Se ha comprobado que la incorporación de estas enzimas (junto con otras hemicelulasas) a su dieta, ha permitido reducir la viscosidad intestinal y optimizar la ganancia nutricional y de peso de los mismos (Kulkarni *et al.*, 1999).

1.6. Nuevas aplicaciones de los módulos de unión a carbohidratos

Existen tres motivos por los cuales el estudio de estas proteínas se está incrementando de manera significativa (Shoseyov *et al.*, 2006):

- a) Pueden funcionar de manera aislada ya que tienen plegamientos propios e independientes.
- b) Se pueden unir a matrices de fijación baratas proporcionando excelentes propiedades físicas y químicas.
- c) Se puede controlar la especificidad de su unión.

Algunas de las aplicaciones concretas de los CBM se indican a continuación:

- Numerosos estudios han afirmado la viabilidad de emplear un CBM como etiqueta de afinidad para la inmovilización de enzimas y su procesamiento. En estos procesos, los carbohidratos se emplean como soporte de afinidad para la posterior inmovilización de la enzima conservando, de forma paralela, su actividad enzimática. Por ejemplo, la familia CBM9 se puede utilizar como una etiqueta para purificar proteínas en una columna de afinidad a base de celulosa (Kavoosi *et al.*, 2004).
- Se pueden diseñar enzimas con funciones mejoradas de reconocimiento de carbohidratos (Cuskin *et al.*, 2012).
- En colaboración con las GH, la elaboración de bioetanol a partir de lignocelulosa es un proceso en el cual participan los CBM. Estas proteínas son cruciales ya que son las responsables de orientar sus módulos catalíticos a los sitios de degradación para la posterior descomposición de biomasa. Tienen un amplio uso en la mejora de enzimas de bioprocesamiento (Reyes-Ortiz *et al.*, 2013).
- Se pueden inmovilizar células sobre superficies de hidratos de carbono (Simsek *et al.*, 2013; Wang & Chao, 2006).

- Se pueden emplear como sondas moleculares para detectar la presencia de hidratos de carbono en plantas y en mamíferos (Boraston *et al.*, 2006; Herve *et al.*, 2010). Desde el punto de vista biológico, esta es la aplicación más importante para poder estudiar la pared celular vegetal.

1.7. *Paenibacillus barcinonensis*

Los miembros que pertenecen al género *Bacillus* suelen ser componentes saprofitos^z de la microbiota del suelo (Claus & Berkeley, 1986). Muchas especies se conocen porque generan enzimas extracelulares las cuales suelen tener aplicaciones industriales importantes (Outtrup & Jørgensen, 2002). La cepa BP-23^T fue aislada a partir del suelo de un arrozal en el delta del río Ebro (España) después de un cultivo de enriquecimiento a temperatura ambiente en paja de arroz durante dos meses (Blanco & Pastor, 1993). Las colonias que mostraron actividad xilanasas se usaron para posteriores estudios (Blanco *et al.*, 1995). Los análisis mostraron que la cepa BP-23^T representa una nueva especie de *Paenibacillus* y se propuso el nombre de *Paenibacillus barcinonensis*^{aa} (Sanchez *et al.*, 2005), la cuál ha demostrado tener un sistema muy eficaz de degradación de xilano.

Hasta la fecha, se han clonado y caracterizado cuatro de sus xilanasas: Xyn10C, Xyn10B, Xyn10A y Xyn30D (Blanco *et al.*, 1999; Gallardo *et al.*, 2010; Valenzuela *et al.*, 2010; Valenzuela *et al.*, 2012). Las tres primeras pertenecen a la familia GH10 y la última a la GH30. La estructura cristalográfica de la xilanasas Xyn10B ha sido anteriormente reportada (Gallardo *et al.*, 2010). El presente trabajo se centra en el estudio de las xilanasas multidominio Xyn10C (CBM22-1-CBM22-2-GH10-CBM9a-CBM9b) y Xyn30D (GH30-CBM35). El objetivo fundamental de la presente tesis ha sido analizar en detalle los determinantes moleculares de reconocimiento proteína-carbohidrato y avanzar en el entendimiento de los sofisticados mecanismos concertados de regulación de la especificidad por los dominios adicionales no catalíticos.

^z Saprofito: adj. biol. Planta o microorganismo que se alimenta de materia orgánica en descomposición.

^{aa} *P. barcinonensis*: adj. Barcinonensis de *Barcino*, el nombre romano de Barcelona, la ciudad en España donde se aisló la cepa.

2. OBJETIVOS

*“Lo importante es no dejar de hacerse preguntas”
“Toda la ciencia no es más que un refinamiento del pensamiento cotidiano”*

Albert Einstein (Alemania, 1879-Estados Unidos, 1955)
Físico. Premio Nobel de Física, 1921

2. Objetivos

El objetivo general del presente trabajo ha sido investigar en detalle los determinantes moleculares de reconocimiento proteína-carbohidrato de dos xilanasas modulares de *P. barcinonensis* así como avanzar en el conocimiento de los mecanismos concertados de regulación de la especificidad por los dominios adicionales no catalíticos, en el contexto general del proceso de degradación de la pared celular vegetal.

El estudio estructural y funcional de estas dos enzimas puede esclarecer los procesos biológicos en los que intervienen. De este modo, los objetivos concretos para cada proteína han sido los siguientes:

Xyn10C:

1. Clonación, expresión, purificación, cristalización y determinación estructural por cristalografía de rayos X del dominio N-terminal Xyn10C-XBD, y de sus módulos aislados CBM22-1 y CBM22-2.
2. Expresión, purificación, cristalización y determinación estructural por cristalografía de rayos X del dominio catalítico Xyn10C-CD.
3. Expresión y purificación de la proteína completa Xyn10C.
4. Realización de ensayos de unión con sustratos de interés mediante ITC y geles de afinidad de cada proteína.
5. Análisis estructural en presencia y ausencia de sustratos específicos.
6. Evaluación de los resultados estructurales y bioquímicos y determinación de la funcionalidad de cada dominio de Xyn10C en la actividad de la enzima.
7. Valoración del papel de la multimodularidad en el mecanismo de degradación de la pared celular vegetal.

Xyn30D:

8. Purificación, cristalización y determinación estructural por cristalografía de rayos X de Xyn30D.
9. Purificación, cristalización y determinación estructural por cristalografía de rayos X de Xyn30D y Xyn30D-CD inactivos así como de sus dominios aislados, Xyn30D-CD y Xyn30D-CBM35.
10. Análisis estructural en presencia y ausencia de sustratos de interés de cada proteína.
11. Realización de ensayos de unión de Xyn30D-CBM35 con sustratos específicos mediante geles de afinidad.
12. Estimación de la contribución de Xyn30D-CBM35 en la despolimerización del xilano.

3. METODOLOGÍA EXPERIMENTAL

*“Fui capturada para la vida por la química y los cristales”
“Una gran ventaja del análisis de rayos X como método de análisis de la
estructura química es su poder para mostrar una estructura totalmente
inesperada y sorprendente con, al mismo tiempo, la certeza completa”*

Dorothy Hodgkin (El Cairo, 1910-Reino Unido, 1994)
Química. Premio Nobel de Química, 1964

3. Metodología experimental

3.1. Metodología en cristalografía macromolecular

3.1.1. Aspectos generales

La inmersión de la química en la biología a principios del siglo XX fue el inicio del estudio estructural del material biológico. Aún así, no fue hasta 1958 cuando se resolvió la primera estructura de una proteína, la mioglobina (Kendrew *et al.*, 1958). Por su parte, Watson y Crick interpretaron de forma correcta datos cristalográficos conseguidos por Rosalind Franklin y, a diferencia de ella, pasaron a la historia por descubrir la doble hélice de ADN (Watson & Crick, 1953).

A lo largo de los años, la cristalografía de rayos X se ha convertido en una de las técnicas físicas más potentes para la resolución estructural de macromoléculas. Presenta grandes ventajas en comparación con otras técnicas enfocadas al mismo fin, ya que proporciona información a escala atómica, obteniendo una “fotografía tridimensional” del material cristalizado. En lo que se refiere a las proteínas, se obtiene su plegamiento y estructura tridimensional, así como una importante información físico-química como distancias y ángulos de enlace, empaquetamiento de los átomos y estado térmico vibracional; datos todos ellos fundamentales para la comprensión de su función biológica.

La técnica alternativa que usa la biología estructural para alta resolución de macromoléculas en disolución es la RMN. Sin embargo, esta técnica está limitada por el tamaño molecular ya que el campo de aplicación de la RMN es para proteínas menores de unos 250-300 residuos. Adicionalmente, otras técnicas como la microscopía electrónica o la microscopía de fuerza atómica ofrecen alternativas al estudio estructural de macromoléculas, pero proporcionan imágenes del contorno general de la proteína a baja resolución. Por último, las técnicas bioinformáticas también pueden descifrar información estructural de una proteína si se conocen datos de una proteína similar, pero nunca aporta la precisión y fiabilidad de la difracción de rayos X. Merece la pena mencionar el reciente descubrimiento introducido en el campo de la nanocristalografía de rayos X, en la escala de tiempo de femtosegundos. La técnica de láser de rayos X de electrones libres (XFEL: *X-ray free electron laser*) consiste en un láser de electrones libres y con longitudes de onda muy reducidas que produce pulsos de rayos X extremadamente brillantes, aproximadamente 1 000 millones de veces superior a la producida actualmente en los sincrotrones. Aún se están dando los primeros pasos en la implantación de esta técnica, pero se prevé que suponga un salto cualitativo en el campo de la biología estructural ya que permitiría prescindir del proceso de cristalización, muchas veces la etapa limitante de la cristalografía (Chapman *et al.*, 2011).

Desde 1958 hasta la actualidad, el incremento de estructuras de macromoléculas determinadas mediante cristalografía de rayos X ha crecido

exponencialmente hasta llegar a casi 100 mil estructuras en 2015 (cifras de la base de datos PDB^{bb}), dentro de las cuales alrededor de 93 mil corresponden a proteínas. El número total de moléculas que forman parte de esta base de datos PDB es de aproximadamente 111 mil. Esta cifra indica que cerca del 90 % de las estructuras depositadas se han resuelto por cristalografía de rayos X.

Esta colección de estructuras, sin embargo, representa solo una pequeña parte de los cientos de miles de diferentes proteínas que desempeñan algún papel en los procesos de la vida. De este modo, la determinación estructural de proteínas es una parte esencial de la investigación que se desarrolla en cientos de laboratorios de todo el mundo.

La técnica de resolución estructural de una macromolécula biológica por cristalografía de rayos X requiere de una serie de etapas que se muestran en el siguiente esquema:

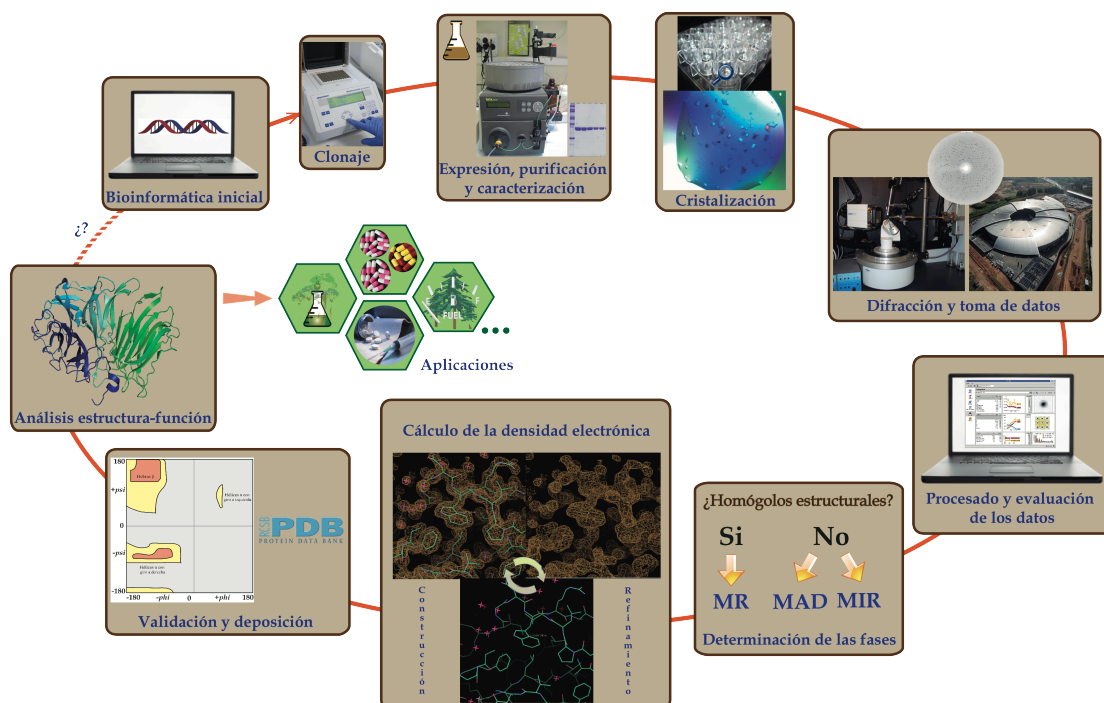


Figura 3.1. Esquema de las distintas etapas necesarias para el estudio estructural mediante cristalografía de rayos X.

3.1.2. Cristalización de proteínas

Para aplicar la técnica de rayos X a una macromolécula, es necesario obtener cristales de la misma. Estos cristales deben ser suficientemente grandes, únicos, homogéneos y de buena calidad para los experimentos de difracción. La etapa de cristalización se ha convertido en el principal obstáculo de la

^{bb} Página electrónica de la base de datos PDB: <http://www.rcsb.org>.

cristalografía macromolecular. No hay una técnica de cristalización que asegure el éxito en este proceso.

3.1.2.1. Concepto de cristal

Se define **cristal** como un sólido constituido por una agrupación ordenada de átomos, iones o moléculas que están repetida y regularmente distribuidos en el espacio. A esta unidad mínima de repetición se le denomina **celdilla unidad** o **elemental**. La celdilla unidad está definida mediante tres ejes (a , b , c) y tres ángulos (α , β , γ), los cuales determinan el sistema de referencia de las coordenadas atómicas. Asimismo, existen otras operaciones de repetición, además de las repeticiones por traslación, que provocan repeticiones dentro de la celdilla elemental: son los **elementos de simetría**. Se denomina **unidad asimétrica** al conjunto mínimo de átomos que, aplicando las operaciones de simetría del grupo espacial, genera el contenido total de átomos de la celdilla (Figura 3.2).

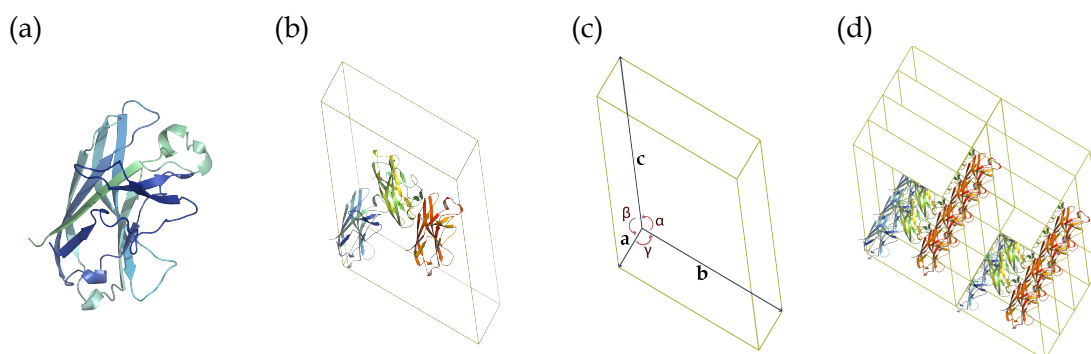


Figura 3.2. La unidad asimétrica (a) se repite a sí misma mediante los elementos de simetría, generando la celdilla elemental (b). Los parámetros que caracterizan su forma y tamaño son tres aristas (a , b , c) y tres ángulos (α , β , γ) (c). La repetición de la celdilla elemental, en las tres direcciones del espacio, genera el cristal (d).

El conjunto de los elementos de simetría de un cristal y los diferentes tipos de celdilla posibles dan lugar a los **grupos espaciales**, los cuales representan 230 maneras distintas en las que un motivo puede repetirse en un espacio tridimensional (Anexo III). Estos 230 grupos están clasificados en las Tablas Internacionales de Cristalografía (*International Tables for X-ray Crystallography*, (Hahn *et al.*, 1984)). De estos 230 grupos, las macromoléculas biológicas son solo compatibles con 65 de ellos. Esto es debido a la quiralidad intrínseca de estas moléculas, lo que impide que formen parte de los grupos de simetría que presenten planos o centros de inversión.

Es importante resaltar que, debido a la naturaleza propia de la biomolécula, los cristales de proteína presentan ciertas particularidades en comparación con cristales inorgánicos y de moléculas pequeñas. Una diferencia notable es que los cristales de proteína suelen ser más pequeños y frágiles. Además, presentan propiedades ópticas más débiles y se deshidratan con facilidad. Esto se debe al alto contenido en solvente que contienen (entre el 25-90 %, con un promedio de

50 %). Esto, *a priori*, puede parecer un problema, debido a que ese alto porcentaje de solvente, en su mayor parte desordenado, se traduce en un patrón de difracción débil. Sin embargo, la presencia del solvente mantiene la proteína en un entorno similar al encontrado en solución. En consecuencia, sus propiedades son las esperadas para la macromolécula en su contexto biológico (McPherson, 2004).

3.1.2.2. Principios básicos de cristalización

La cristalización de una macromolécula no es, en principio, muy diferente a la cristalización de una molécula pequeña. Los cristales de cualquier sustancia se forman a partir de una disolución sobresaturada, lo cual desencadena la formación espontánea del núcleo inicial de cristalización. Cuando se supera el límite de solubilidad (sobresaturación) las moléculas que están en exceso se separan de la fase acuosa y pasan a la fase sólida, estableciéndose el equilibrio entre ambas fases. Por lo tanto, las condiciones de sobresaturación son necesarias pues si hubiera únicamente saturación no se produciría el cambio de fase y, por lo tanto, no se formaría el cristal (Figura 3.3). La estrategia, por consiguiente, es obtener una solución sobresaturada que genere unos pocos núcleos de cristalización sobre los que crezcan unos pocos cristales grandes y ordenados. Es decir, se debe mantener la proteína cerca de la curva de solubilidad, pero en la zona de sobresaturación.

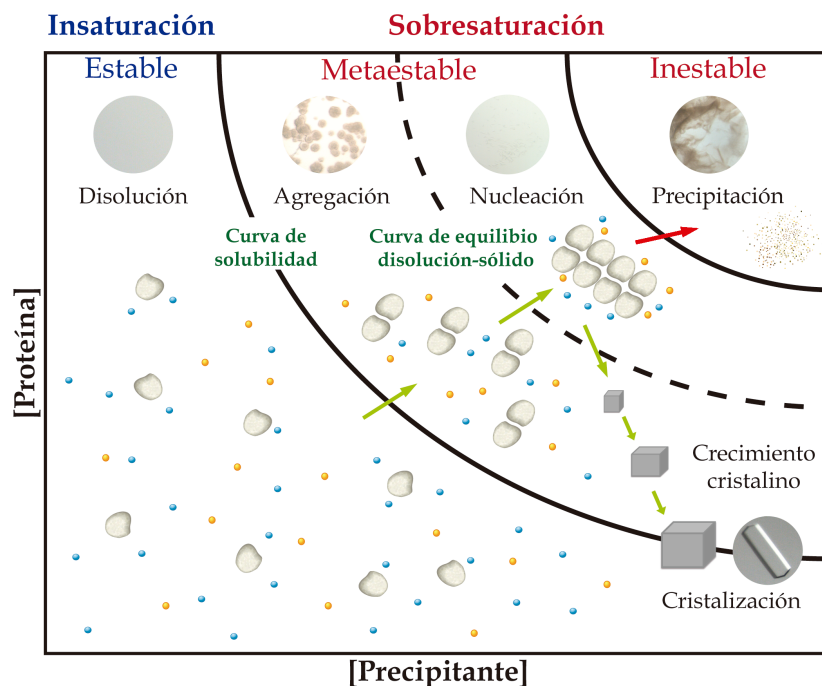


Figura 3.3. Diagrama de fases del proceso de cristalización. Se representan las fases acuosa y saturada además de las zonas de agregación, nucleación, crecimiento cristalino y precipitación. Las flechas verdes muestran la evolución habitual de un experimento de difusión de vapor hasta la formación de cristales. Si se formara precipitado, seguiría el camino de la flecha roja. Los círculos azules representan el solvente y los naranjas el precipitante.

La cristalización consta de dos etapas bien diferenciadas: nucleación y crecimiento cristalino. Por un lado, en la nucleación aparecen agregados cristalinos termodinámicamente estables. Por otro lado, en el crecimiento las moléculas de proteína se van incorporando a la superficie del cristal, desplazando parte de las moléculas de agua. Esta fase acaba cuando se logra el equilibrio termodinámico o cuando se acumulan imperfecciones en el cristal (maclas). La nucleación puede ocurrir de forma espontánea o se puede inducir artificialmente. Por lo tanto, el estudio de la nucleación es una parte fundamental en el crecimiento cristalino. Varias teorías se han propuesto para explicar este fenómeno, entre ellas, la ecuación de Gibbs-Thomson es elemental para explicar los eventos de nucleación (Dhanaraj *et al.*, 2010).

3.1.2.3. Requisitos previos a la cristalización

Para que la cristalización sea posible, lo primero que hay que considerar es si la muestra de la que disponemos tiene los requerimientos necesarios. Existen cinco aspectos principales a tener en cuenta: concentración, pureza, homogeneidad, estabilidad y cantidad (Bergfors, 2009).

- Concentración: Debido a que se necesita alcanzar la sobresaturación, se requiere inicialmente una concentración elevada de proteína. Se consideran adecuadas concentraciones del orden de 5-20 mg/ml.
- Pureza: Se requiere la máxima posible, aunque niveles del 95 % son aceptables. Esta se suele determinar mediante electroforesis en gel de poliacrilamida. Las impurezas pueden detener el crecimiento del cristal y producir irregularidades en el mismo, lo que se traduce en una peor calidad del patrón de difracción.
- Homogeneidad: La muestra debe ser homogénea y cualquier factor que induzca heterogeneidad debe ser evaluado y considerado antes del experimento de cristalización.
- Estabilidad: Hay que tener cuidado de no alterar la estabilidad de la proteína en los procesos de purificación, congelación/descongelación y cristalización. Esto puede ser perjudicial para la formación del cristal.
- Cantidad: Es importante tener una cantidad aceptable de proteína que permita realizar la máxima cantidad de ensayos de cristalización posibles. No es recomendable usar muestra de proteína de diferentes lotes de purificación, ya que esta variable puede ser clave en el proceso de cristalización.

Una vez que la proteína cumple todos estos requisitos, la estrategia más utilizada para alcanzar la sobresaturación se basa en aumentar la concentración por eliminación de solvente y añadir a la solución concentrada de proteína sustancias que disminuyen su solubilidad. Habitualmente, los factores que afectan en la obtención de soluciones sobresaturadas de proteína, son los siguientes (Bergfors, 2009): concentración de proteína, pH y tampón de la solución, temperatura, fuerza iónica, tipo y concentración de agente precipitante, aditivos y ligandos.

3.1.2.4. Experimentos de cristalización

3.1.2.4.1. Metodologías de cristalización

Existen varias técnicas de cristalización de proteínas: difusión de vapor, diálisis, *microbatch* bajo aceite y contradifusión. La difusión de vapor es la técnica más usada y la que se ha empleado en esta tesis en sus dos modalidades: gota sentada y gota colgante.

Difusión de vapor

En este sistema, una gota de solución de proteína junto con otra de agente precipitante se sitúan en un recipiente cerrado y se equilibran frente a la solución precipitante contenida en un reservorio. De este modo, al ser la concentración de precipitante menor en la gota, se produce la evaporación, disminuyendo el volumen y aumentando la concentración de proteína hasta llegar al equilibrio. Así, primeramente se alcanzarán las condiciones de saturación y después de sobresaturación, llegando a la zona de nucleación. Conociendo el sistema y ajustando los parámetros oportunos en diferentes experimentos, el objetivo es llegar a la zona de nucleación y crecimiento cristalino.

Existen dos modalidades experimentales para el montaje de este sistema (Figura 3.4):

- **Gota sentada:** La gota con la proteína y el precipitante se colocan sobre una elevación plana o cóncava que rodea el reservorio de solución precipitante, situada por encima del nivel de dicha solución. El sistema se sella del exterior con una lámina adhesiva flexible que cubre todo el pocillo.
- **Gota colgante:** La gota con la proteína y el precipitante se colocan sobre un cubreobjetos, el cual está siliconizado para evitar la dispersión. La disolución de precipitante se deposita en un pocillo sobre el cual se invierte el cubreobjetos con la gota adherida y se utiliza para cerrar el pocillo utilizando grasa de vacío para sellarlo herméticamente.

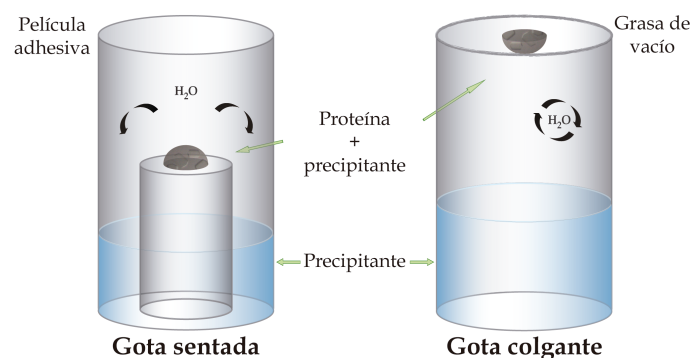


Figura 3.4. Esquema de la técnica de cristalización de difusión de vapor en gota sentada y gota colgante.

La cristalización implica dos etapas experimentales principales: **búsqueda de condiciones de cristalización** y **optimización** de las mismas. Las cuales se describen en los apartados siguientes.

3.1.2.4.2. *Búsqueda de condiciones de cristalización*

Cuando se inician los experimentos de cristalización de una proteína, es necesario realizar una gran cantidad de pruebas preliminares que permitan ensayar el máximo de entornos posibles. Existen múltiples variables que afectan a la cristalización, por lo que hay que tener en cuenta factores químicos (pH, fuerza iónica, concentración, impurezas, etc.), físicos (temperatura, presión, vibración, etc.) y bioquímicos (pureza, mutaciones, ligandos, etc.).

Las técnicas de alto rendimiento nos permiten realizar numerosos experimentos en poco tiempo y gastar la mínima cantidad de muestra posible. Para ello, existe una amplia variedad de kits comerciales que permiten analizar, en un solo experimento, casi cien condiciones diferentes de precipitación. Normalmente, las soluciones de precipitación suelen contener un tampón que mantiene el pH y uno o varios agentes precipitantes. En los últimos años, la aparición de robots y granjas de almacenamiento han permitido automatizar y monitorizar la búsqueda de condiciones, lo cual ha aumentado la efectividad y rapidez y, con ello, el número de biomoléculas cristalizadas.

Los robots de cristalización utilizados en este trabajo han sido *Gilson Quad-Z 215 Liquid Handler* e *Innovadine Technologies Inc. Nanodrop™ ExtY*, (Figura 3.5) y los paquetes de soluciones comerciales: PCT, Crystal Screen , Crystal Screen 2, Index y SaltRx de la casa comercial *Hampton Research* y JCSG+Suite y PACT Suite de *Qiagen*.

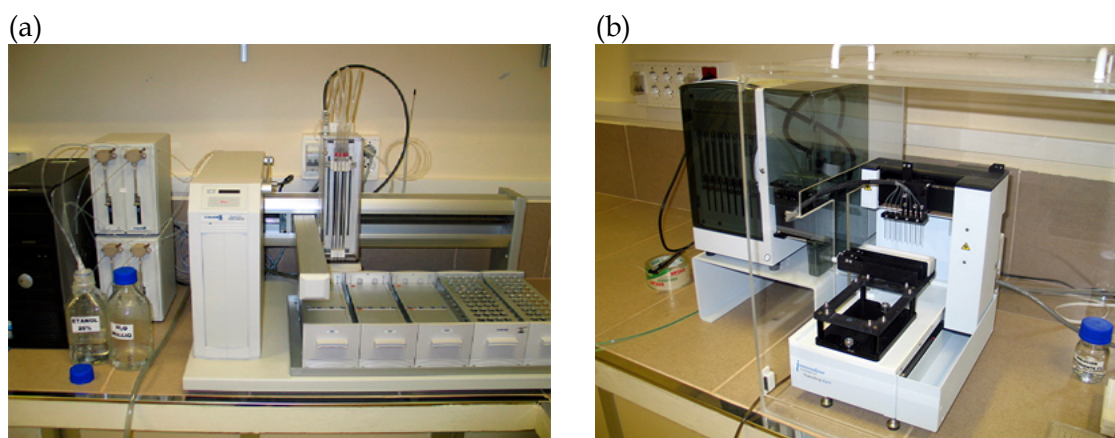


Figura 3.5. Equipos utilizados en el trabajo. (a) Robot *Quad-Z 215 Liquid Handler*. Se emplea para preparar las cajas llenando los pocillos con las soluciones precipitantes a investigar. (b) Robot *Innovadine Technologies Inc. Nanodrop™ ExtY*. Su función es dispensar las gotas de precipitante y proteína.

3.1.2.4.3. Optimización de los cristales

Habitualmente, los cristales que se obtienen en esa primera etapa suelen ser muy pequeños y muchos presentan imperfecciones, por lo que es necesario mejorar su calidad. Para ello, la práctica más común es el **escalado**. Esta técnica consiste en aumentar los volúmenes de trabajo con la muestra de proteína y precipitante. El objetivo es ampliar el tamaño de los cristales y su grado de orden y, con ello, intentar mejorar la calidad del patrón de difracción.

A la vez que se intenta reproducir los cristales, se suelen introducir pequeños cambios dentro de la solución de cristalización (pH, temperatura, concentración de precipitante, etc.), así como también se puede variar la proporción de los componentes de la gota (proteína/precipitante; 1:2, 2:1, 1:3, 3:1, etc.) e incluso introducir pequeñas cantidades de aditivos o ligandos propios de la macromolécula con el fin de mejorar la calidad de los cristales.

Cuando la utilización de las estrategias anteriores no es efectiva, se suele recurrir a una técnica que en la presente tesis ha dado resultados muy favorables: el “**sembrado**” (*seeding*) (Bergfors, 2003). Esta técnica consiste en disminuir la sobresaturación del sistema con el objetivo de situarlo en la zona metaestable dentro de la curva de solubilidad. Después, se añaden núcleos o micro cristales formados previamente, de tal forma que la proteína presente en la solución pueda ir incorporándose a los mismos generando así cristales de mayor tamaño.

Existen tres variantes dentro de esta técnica de sembrado: “micro-sembrado” (*microseeding*), “macro-sembrado” (*macroseeding*) y “sembrado por arrastre” (*streak-seeding*). El “micro-sembrado” consiste en triturar un cristal defectuoso obtenido previamente, y usar sus núcleos como base para el crecimiento de nuevos cristales (Luft & DeTitta, 1999). El “macro-sembrado” se basa en manipular un cristal de tamaño aceptable pero con imperfecciones. Para ello se realizan lavados superficiales del mismo y se usa como núcleo de cristalización. Por último, el “sembrado por arrastre” consiste en introducir en una solución con núcleos cristalinos o pequeños cristales una fibra natural y arrastrar los núcleos adheridos a ella a la solución de interés (Figura 3.6). Esta es la técnica de sembrado que hemos utilizado en el trabajo.

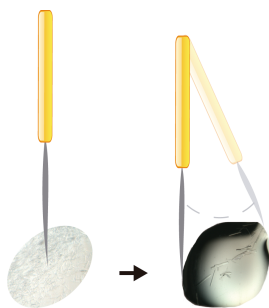


Figura 3.6. Esquema de la realización de un experimento de “sembrado por arrastre” (*streak-seeding*).

3.1.2.4.4. *Cristalización de complejos proteína-ligando.*

Conocer la estructura tridimensional de la proteína en presencia de ligandos que interaccionen con ella, como pueden ser sustratos o productos de reacción, es clave para conocer el mecanismo de acción de la biomolécula. Para la cristalización de complejos proteína-ligando existen dos estrategias: inmersión (*soaking*) o cocrystalización (Hassell *et al.*, 2007).

- **Inmersión** (*soaking*): Los cristales de proteína se introducen a una disolución que contiene el ligando a alta concentración. El ligando puede difundir hacia el sitio activo o de unión debido a los canales que existen dentro del cristal. La desventaja que presenta esta estrategia es que muchas veces los cristales se dañan con la entrada del sustrato o, simplemente, el sustrato no puede acceder por impedimento del empaquetamiento cristalino.
- **Cocrystalización**: En esta estrategia, el ligando se añade a la gota de proteína al comenzar el experimento de cristalización o bien se incubaba con ella previamente. Generalmente, suele usarse la misma condición de cristalización que para la proteína silvestre pero muchas veces (debido a cambios en la conformación, por ejemplo), se debe buscar una nueva condición de cristalización.

3.1.2.5. *Criocristalografía*

Para obtener datos completos de la muestra, se tiene que conseguir un mapa de difracción completo del cristal. Debido a que la radiación daña los cristales, esto resulta muy complicado a temperatura ambiente. Tradicionalmente, la estrategia había sido realizar recogidas de datos de diferentes cristales, que después se combinaban para llegar a resultados aceptables, pero con muchas limitaciones. La introducción de las técnicas de criocristalografía ha resuelto este problema (Alcorn & Juers, 2010). Los experimentos se realizan a temperaturas de unos 100 K y, de esta manera, la muestra cristalina es capaz de resistir haces intensos de luz, como los producidos por radiación sincrotrón.

Una de las limitaciones de esta técnica es el riesgo de la formación de cristales de hielo procedentes del solvente, lo que afecta al patrón de difracción y puede incluso destruir el cristal. Para evitar esto, debe tratarse la muestra con soluciones crioprotectoras. De esta manera, el crioprotector elimina las moléculas de agua no estructurales para evitar la formación de cristales de hielo (Rubinson *et al.*, 2000). Según los componentes de la solución de cristalización, se sigue una estrategia adecuada en la elección y preparación de la solución de crioprotección. Los crioprotectores más comunes suelen ser efectivos a una concentración entre 15-25 % (*v/v*) y se preparan en la solución madre de cristalización. Los más comunes son glicerol (Kempkes *et al.*, 2008), etilenglicol, aceite mineral y PEG 400, pero también son efectivas algunas sales como por

ejemplo malonato (Holyoak *et al.*, 2003), agente incluido en una de las condiciones de cristalización de las proteínas en estudio en este trabajo.

El proceso de crioprotección también resulta ventajoso para el transporte de los cristales. Estos se colocan en un lazo de nylon, que es transparente a los rayos X, y quedan depositados sobre la fina película que forma la criosolución debido a la tensión superficial (Figura 3.7a). El lazo está unido a la cara interna del tapón magnético de un vial, el cual va relleno de nitrógeno líquido durante el transporte de las muestras (Figura 3.7b). Finalmente, el tapón se adhiere a la cabeza goniométrica (Apartado 3.1.3.1.2.) del equipo de difracción de forma muy rápida y eficaz (Figura 3.12). Los viales suelen transportarse en dispositivos de 10-16 muestras (Figura 3.7c) y éstos sumergirse en un contenedor enfriado previamente con nitrógeno líquido (Figura 3.7d).

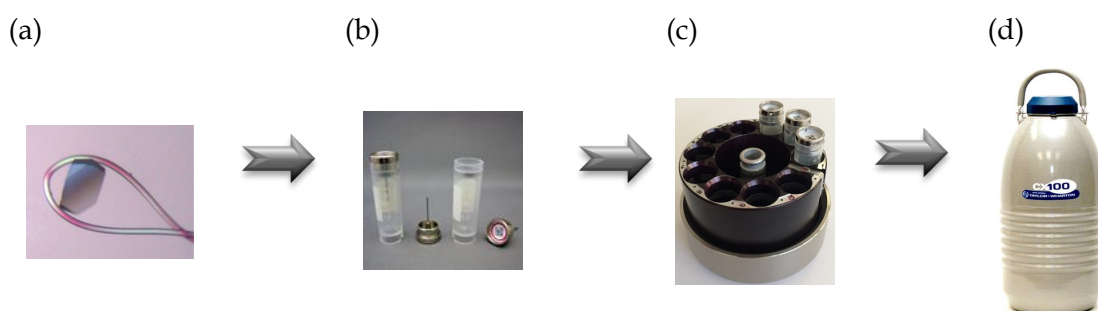


Figura 3.7. Esquema del procedimiento de montaje y traslado del cristal.

3.1.3. Difracción de rayos X

3.1.3.1. Introducción al fenómeno de difracción

La difracción de rayos X es el resultado de la interacción de la radiación electromagnética con los electrones de los átomos en el cristal. Esta interacción produce una vibración acoplada de los electrones con la variación del campo eléctrico. Así, parte de los rayos atraviesan el cristal y parte son dispersados por los electrones del mismo. Estos fotones dispersados pueden anularse (si sus ondas interfieren) o, debido a su naturaleza periódica, pueden emerger en fase y reforzarse, generando el espectro de difracción de rayos X. Este fenómeno se puede explicar geométricamente mediante la ley de Bragg (Ecuación 3.1) (William L. Bragg, Premio Nobel de Física en 1915):

$$2d_{hkl}\sin\theta = n\lambda \quad (\text{Ecuación 3.1})$$

En esta expresión se consideran diferentes planos atómicos paralelos (hkl) separados por una distancia (d), con un ángulo de incidencia del haz de rayos X (θ) y una longitud de onda del haz (λ) (Figura 3.8).

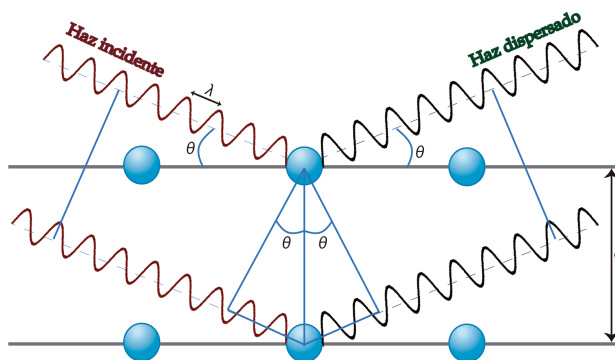


Figura 3.8. Ley de Bragg. La onda que incide en el plano inferior se retrasa un número entero de longitudes de onda y sale en fase con la onda superior. De esta manera, se refuerza la señal.

Para explicar mejor el fenómeno de la difracción, se usa el concepto de red recíproca y el modelo de Ewald. En este modelo, se reemplaza cada familia de planos de Bragg por un punto geométrico (Figura 3.9), de manera que todas las familias de planos constituyen una red recíproca, íntimamente asociada con la red cristalina. De este modo, al girar el cristal la difracción ocurrirá cuando cada uno de los puntos de la red recíproca interaccione con la superficie de la esfera de Ewald (cuyo radio es igual al inverso de la longitud de onda). Finalmente, situando el detector en la posición adecuada, se puede recoger el haz difractado.

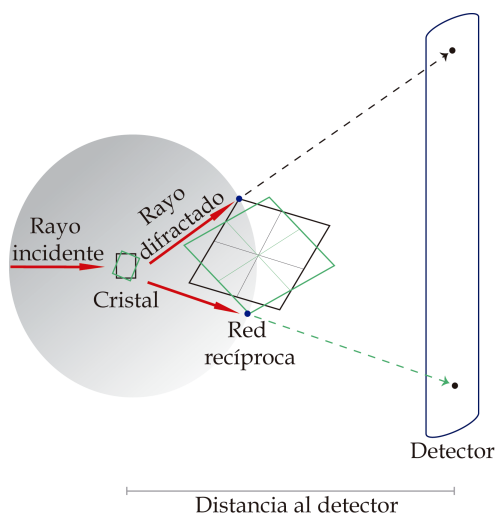


Figura 3.9. Representación de la esfera de Ewald. Al encontrarse un punto recíproco (azul) con la esfera de Ewald se cumple la ley de Bragg.

Para realizar el experimento de difracción se requiere de fuente de rayos X, goniómetro, crioprotección y detector.

3.1.3.1.1. Fuentes de radiación de rayos X

La historia de los rayos X comenzó con William Crookes en el siglo XIX. Este fisicoquímico estaba estudiando la generación de corrientes de alto voltaje en un tubo de vacío cuando vio que si este estaba cerca de películas fotográficas, se generaban imágenes borrosas en ellas. A partir de ahí, Wilhelm

Röntgen siguió con los estudios de Crookes y descubrió estos desconocidos (hasta ese momento) rayos X. Con ello ganó el Premio Nobel de Física en 1901.

Los rayos X son ondas electromagnéticas extraordinariamente penetrantes que atraviesan ciertos cuerpos, producidas por la emisión de los electrones internos del átomo. De esta manera, son generados por la desaceleración rápida de electrones muy energéticos al chocar con un blanco metálico. Normalmente, la longitud de onda mas usada en cristalografía de proteínas es del orden de 1 Å, adecuada para distancias atómicas. Las fuentes más utilizadas son los generadores de ánodo rotatorio y las estaciones de radiación sincrotrón. En el trabajo se han utilizado ambas fuentes, principalmente la radiación sincrotrón.

- **Generadores de ánodo rotatorio:** Los difractómetros convencionales que se suelen encontrar en los laboratorios de cristalografía de molécula pequeña contienen tubos de alto voltaje, en cuyo extremo se ubica una placa metálica (generalmente de cobre), la cual es bombardeada con electrones acelerados generados a partir de un filamento incandescente (Figura 3.10a). Cuando estos electrones excitados vuelven al estado fundamental, se emiten los rayos X con una frecuencia proporcional al salto de energía que provocó su excitación.

Este proceso genera mucho calor y por ello se introdujeron los generadores de ánodo rotatorio, en los cuales el ánodo metálico es un cilindro que gira para conseguir una mejor refrigeración, lo que permite obtener rayos X mucho más intensos que a partir de un tubo convencional (Figura 3.10b).

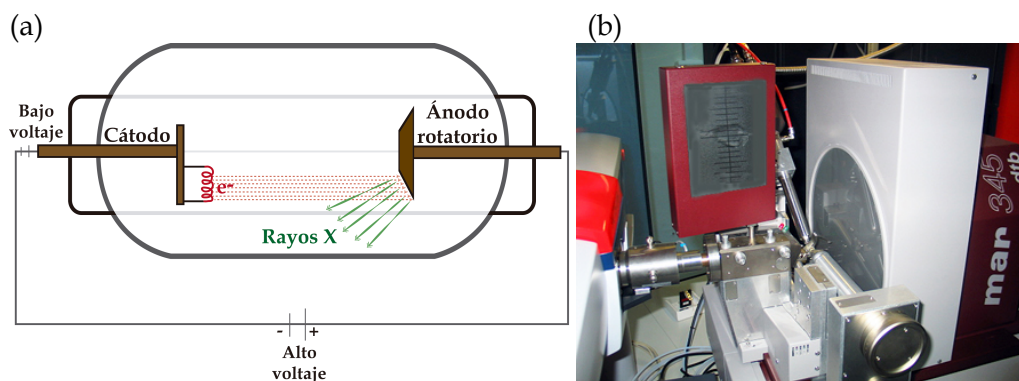


Figura 3.10. (a) Representación esquemática de un tubo de rayos X en ánodo rotatorio. (b) Generador de rayos X (*MicroStar* 2,7 kW, *Bruker*) de ánodo rotatorio y detector (*MAR345dtb*, *MarResearch*) perteneciente al Instituto de Química Física "Rocasolano", CSIC, España.

- **Radiación sincrotrón:** Un sincrotrón es un acelerador de partículas (generalmente electrones) en el que éstas se desplazan en una órbita circular por la acción de un campo magnético y son aceleradas a velocidades cercanas a la luz por un campo eléctrico sincronizado con el movimiento orbital (Figura 3.11a). En el momento en el cual se rompe la trayectoria

rectilínea, los electrones emiten una radiación muy energética: La radiación sincrotrón, la cual engloba distintas longitudes de onda, desde las microondas a los rayos X.

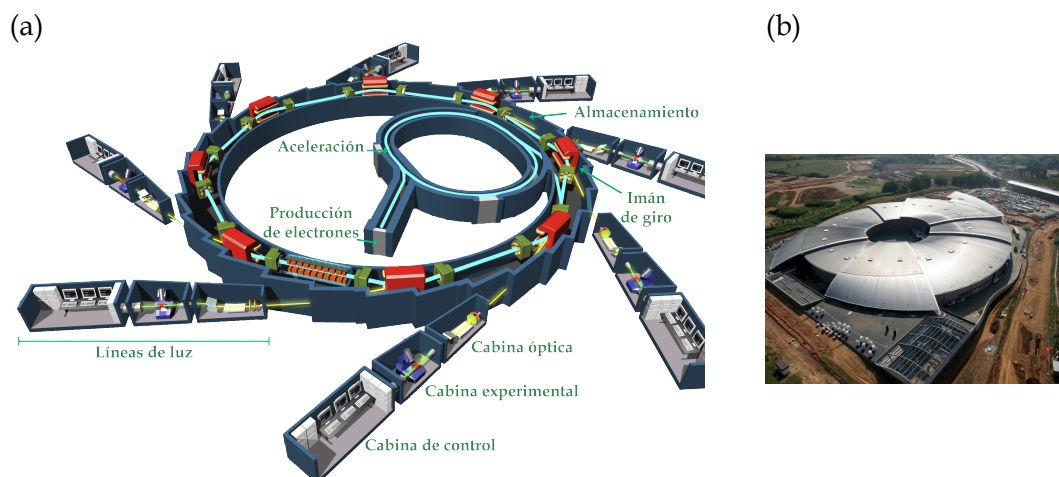


Figura 3.11. (a) Esquema general del funcionamiento de un sincrotrón. (b) Imagen exterior del sincrotrón ALBA (Barcelona, España). Figuras adaptadas de www.cells.es.

Las estaciones de radiación sincrotrón en las cuales se han realizado recogidas de datos de difracción para la presente tesis son los siguientes: **ESRF^{cc}** (Grenoble, Francia), **ALBA** (Barcelona, España), **DESY^{dd}** (Hamburgo, Alemania) y **Diamond** (Londres, Reino Unido).

3.1.3.1.2. Crioprotección

Los átomos de proteína y solvente pueden sufrir daños irreparables debido a la radiación, ya que esta genera calor y radicales libres que degradan el cristal (Garman & Owen, 2006). Para solventarlo, se utilizan sistemas de baja temperatura para enfriar el cristal con una corriente de nitrógeno líquido (Figura 3.12) alcanzando temperaturas de 100 K. Este tratamiento puede aumentar la vida media del cristal y su resolución ya que se fijan zonas flexibles en la proteína gracias a la disminución de su vibración térmica (Garman, 1999). Para trabajar a estas temperaturas, previamente debe crioprotgerse el cristal como se ha visto en el Apartado 3.1.2.5.

Después de la correcta crioprotección del cristal, éste se captura con un lazo de nylon y se sitúa en un aparato que recibe el nombre de **goniómetro**. La cabeza goniométrica es un dispositivo que posee varios grados de libertad y permite orientar el cristal en intervalos pequeños del orden de $0,2^{\circ}$ - 1° , mientras inciden sobre él los rayos X. De esta forma, se puede girar el cristal para conseguir un conjunto de datos completo de difracción.

^{cc} ESRF: Instalación europea de radiación sincrotrón (*European synchrotron radiation facility*).

^{dd} DESY: Sincrotrón alemán de electrones (*Deutsches elektronen synchrotron*).

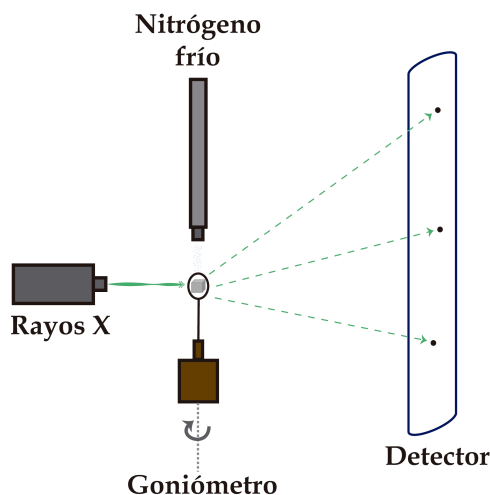


Figura 3.12. Esquema del montaje y colección de datos de un cristal.

3.1.3.1.3. Detector

Actualmente, casi todos los detectores utilizados en cristalografía macromolecular son detectores de área. Los tres más conocidos son:

- **Detectores de placa de imagen, IP (*Image Plate*):** En este caso existe una placa que contiene una sal de un lantánido. Los átomos se excitan cuando son irradiados con un fotón procedente del haz de rayos X difractado. A continuación, un láser barre la placa lo que hace que se produzca una emisión de luz proporcional al número de fotones absorbidos. Esta luz es registrada por un fotomultiplicador y se convierte en una imagen del patrón de difracción. El proceso dura unos 4-9 minutos y requiere borrar la placa entre cada medida, por lo que su uso está limitado a generadores de ánodo rotatorio (Figura 3.10b).
- **Detectores CCD (*Charged-Coupled Device*):** Se basan en el efecto fotoeléctrico, ocurriendo la conversión de la luz recibida en forma de fotones en corriente eléctrica. Por lo tanto, constan de un sensor con pequeñas células fotoeléctricas que registran la imagen. Estos dispositivos son más sensibles y tienen tiempos de exposición menores que los detectores IP. Suelen emplearse en instalaciones sincrotrón.
- **Detectores de matriz de píxeles, PAD (*Pixel Array Detectors*):** Este dispositivo contiene una capa de silicio muy sensible que absorbe los rayos X y los convierte en una señal eléctrica. El sistema de lectura es extremadamente eficaz y rápido, por lo que están presentes en instalaciones sincrotrón. Otra gran ventaja es que existe muy poco ruido de fondo.

3.1.3.2. Recogida de datos de difracción

En esta fase se colectan una serie de imágenes de difracción consecutivas para distintas posiciones del cristal. El fin es conseguir un conjunto de datos completo. Usar una buena estrategia de toma de datos es fundamental ya que de ello dependerá la calidad de los mismos para el posterior proceso de resolución estructural. Para ello, deben controlarse una serie de parámetros, como son: tiempo de exposición, ángulo de giro por imagen ($\Delta\phi$), ángulo de rotación total, distancia al detector, etc. (Dauter, 1999).

3.1.4. Resolución estructural

3.1.4.1. Procesado de datos

Una vez finalizado el experimento de difracción, se procesan todas las imágenes obtenidas (Figura 3.13) para extraer las intensidades netas de los haces difractados.

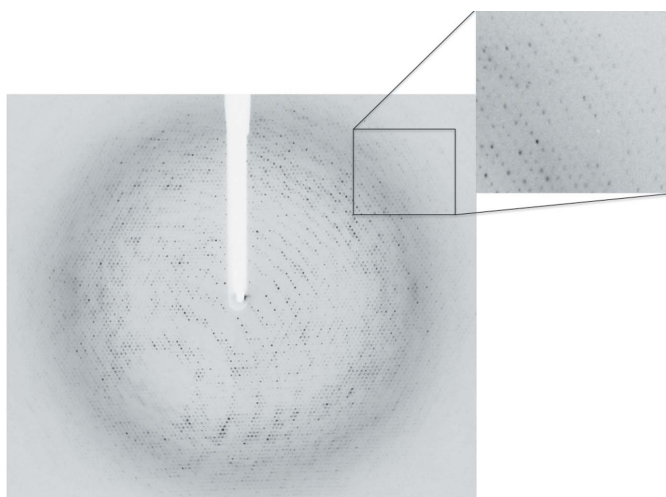


Figura 3.13. Patrón de difracción de la proteína Xyn30D a 2,3Å de resolución. Detalle ampliado del patrón de la zona recuadrada que muestra reflexiones de distinta intensidad.

Este proceso está constituido por las siguientes etapas:

- **Indexado:** La disposición geométrica (posición) de las reflexiones en el patrón de difracción nos permite extraer información de la celdilla unidad de la red cristalina y de la simetría del cristal. De este modo, se asignan los índices de Miller (hkl) a cada una de las reflexiones. Por lo tanto, se determina la celdilla unidad, el grupo espacial y la mosaicidad del cristal^{ee} (Bellamy *et al.*, 2000; Harp *et al.*, 1998).

Para esta etapa se han utilizado los programas **iMosflm** (Battye *et al.*, 2011), **XDS** (Kabsch, 2010) y **Pointless** (Evans, 2006).

^{ee} Mosaicidad: Parámetro que estima el grado de desorden de un cristal.

- **Integrado:** En esta etapa se estima la intensidad de cada reflexión y su error relativo. Por consiguiente, se obtiene el conjunto de todas las reflexiones (hkl) junto con sus intensidades (I_{hkl}) y sus errores asociados (σ_{hkl}).

Para esta etapa se han utilizado los programas **iMosflm** (Battye *et al.*, 2011) y **XDS** (Kabsch, 2010).

- **Escalado:** En esta etapa se escalan y normalizan las intensidades, reduciendo los datos a la unidad asimétrica. Esta normalización es necesaria debido a factores experimentales (temperatura, intensidad de la fuente y disminución de las intensidades por destrucción del orden cristalino). Además, debido a la simetría contenida en el cristal existe una alta redundancia de datos, lo que permite obtener una serie de parámetros estadísticos que dan idea de la calidad de los datos en función de la resolución. Son los siguientes:
 - **Multiplicidad:** Es el número de veces que se ha registrado una determinada reflexión. Cuanto mayor sea su valor, mejor será la estimación de errores asociados a la medida de la reflexión.
 - **Compleitud:** Indica el porcentaje de espacio recíproco que ha sido analizado.
 - $\langle |I|/\sigma(I) \rangle$: Representa la relación entre la intensidad con respecto al ruido de fondo. Son aceptables valores mayores de 2 para la última capa de resolución. Su expresión es (siendo N el número de reflexiones):

$$\langle \frac{|I|}{\sigma(I)} \rangle = \frac{1}{N} \sum_{hkl} \frac{|I_{hkl}|}{\sigma(I_{hkl})} \quad (\text{Ecuación 3.2})$$

- **Factores R_{merge} y R_{pim} :** R_{merge} evalúa la diferencia entre la intensidad de una reflexión medida N veces y su promedio (Ecuación 3.3). Existe un nuevo indicador denominado R_{pim} , el cual tiene en cuenta, además, el número total de reflexiones (Ecuación 3.4). Valores máximos de 0,3-0,4 para la última capa de resolución son aceptables.

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^N |I_{(hkl)i} - \bar{I}_{(hkl)}|}{\sum_{hkl} \sum_{i=1}^N I_{hkl}} \quad (\text{Ecuación 3.3})$$

$$R_{pim} = \frac{\sum_{hkl} \sqrt{\frac{1}{N-1} \sum_{i=1}^N |I_{(hkl)i} - \bar{I}_{(hkl)}|}}{\sum_{hkl} \sum_{i=1}^N I_{hkl}} \quad (\text{Ecuación 3.4})$$

Para esta etapa se utilizaron **Aimless** y **Scala** (Evans, 2011), que forman parte del paquete informático **CCP4** (*Collaborative Computational Project, Number 4*) (Winn *et al.*, 2011).

- **$CC_{1/2}$ y CC^*** : Son dos parámetros basados en el cálculo del coeficiente de correlación de Pearson (CC) (Karplus & Diederichs, 2012). Estos nuevos parámetros consideran que si se divide el conjunto de datos en dos, los errores en cada uno serían aleatorios. Su relación está expresada en la Ecuación 3.5:

$$CC^* = \sqrt{\frac{2CC_{1/2}}{1 + CC_{1/2}}} \quad \text{(Ecuación 3.5)}$$

- **Factor de estructura (F_{hkl})**: Se define como la resultante de la dispersión de los rayos X por todos los átomos de la celdilla en una determinada dirección. Esta resultante es una onda y como tal, tiene un vector asociado que puede describirse por su módulo, $|F_{hkl}|$, y su fase ϕ_{hkl} . Los módulos representan las amplitudes de los haces difractados y se obtienen directamente a partir de las intensidades recogidas. Sin embargo, las fases ϕ_{hkl} no pueden determinarse experimentalmente y se requiere de distintos métodos para su cálculo (como se verá a continuación).

La expresión analítica de los factores de estructura se muestra en la Ecuación 3.6. En ella aparece la magnitud f , que es el factor de dispersión atómico y representa la dispersión de los rayos X producida por los electrones del átomo:

$$F_{hkl} = \sum_j f_j e^{2\pi i(hx_j + ky_j + lz_j)} \quad \text{(Ecuación 3.6)}$$

Para la determinación de F a partir de I , se utilizó el programa **Truncate** (Evans, 2011), que forma parte del paquete informático **CCP4** (Winn *et al.*, 2011).

3.1.4.2. Resolución del problema de la fase

La transformación del espacio recíproco (espectro de difracción) al espacio cristalino (función de densidad electrónica) implica la resolución de una función denominada transformada de Fourier. Esta ecuación relaciona la densidad electrónica, ρ_{xyz} para cada punto de la celdilla elemental cristalina con el espacio recíproco, el cual está definido por los factores de estructura F_{hkl} . Se expresa como (Ecuación 3.7, siendo V el volumen de la celda unidad):

$$\rho(x, y, z) = \frac{1}{V} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} |F_{hkl}| e^{-2\pi i(hx + ky + lz - \phi_{hkl})} \quad \text{(Ecuación 3.7)}$$

Por tanto, para calcular los mapas de densidad electrónica sobre los cuales se va a construir el modelo atómico, es necesario conocer tanto las amplitudes asociadas a cada haz difractado $|F_{hkl}|$ extraídas del patrón de difracción, como su fase ϕ_{hkl} que es desconocida. Esto constituye el llamado “problema de la fase” en cristalografía de proteínas, lo que puede solucionarse mediante tres métodos (Taylor, 2010):

- a) **Método de Reemplazo Molecular (MR, *Molecular Replacement*)**: Se utiliza una proteína homóloga de estructura conocida como modelo estructural inicial.
- b) **Método de Reemplazo Isomorfo Simple/Múltiple (SIR/MIR, *Single/Multiple Isomorphous Replacement*)**: Se difunden en el cristal átomos pesados con elevado poder de dispersión. Mediante la función de Patterson y debido a la diferencia de intensidad entre los espectros de cristales con (derivados) y sin (nativos) átomo pesado, se puede estimar la posición de los átomos dentro de la celdilla. (SIR: un solo cristal; MIR: varios cristales derivados). A partir de las posiciones de estos átomos pesados, se pueden asignar fases a todas las reflexiones del espectro.
- c) **Método de Dispersión Anómala Simple/Múltiple (SAD/MAD, *Single/Multi-wavelength Anomalous Diffraction*)**: Se introducen átomos que se comportan como dispersores anómalos para la longitud de onda incidente, modificando el valor de dispersión atómico (f_j , Ecuación 3.6). (SAD: una única longitud de onda; MAD: diferentes longitudes de onda).

Además, se pueden combinar los métodos de reemplazo isomorfo y dispersión anómala (SIRAS/MIRAS) (Taylor, 2010), técnicas muy potentes para resolver *ab initio* el problema de la fase en cristales de proteínas nuevas.

En el trabajo se ha utilizado el método de MR como técnica de resolución estructural de las proteínas en estudio, y sus fundamentos se describen a continuación.

3.1.4.2.1. Método de Reemplazo Molecular

Este método se basa en el supuesto de que proteínas homólogas en secuencia peptídica tienen un plegamiento similar y, por lo tanto, pueden ser usadas como modelos estructurales. La técnica consiste en situar la macromolécula conocida en el cristal de la proteína desconocida, para lo cual es necesario conocer su orientación y posición concreta. El procedimiento se basa en el uso de la función de Patterson y el cálculo de funciones de rotación y de traslación (Figura 3.14).

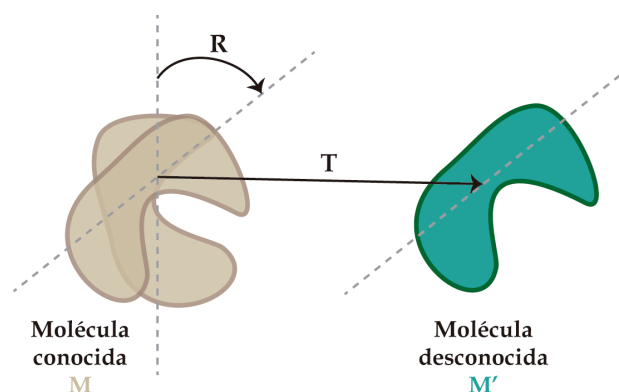


Figura 3.14. Esquema del método de reemplazo molecular. La molécula de estructura conocida y proteína modelo (M) se gira mediante la operación (R) y se traslada mediante la operación (T), para llevarla a la posición de la molécula desconocida (M'). Después debe evaluarse si esta posición es la correcta o si se debe seguir buscando la solución adecuada.

Esta estrategia se usa también para resolver la estructura de complejos proteína-ligando en el caso de que los cristales pertenezcan a distinto grupo espacial que los cristales de la proteína.

En el trabajo se han utilizado los programas de MR **Phaser** (McCoy *et al.*, 2007) y **Molrep** (Vagin & Teplyakov, 2010), ambos parte del paquete informático **CCP4** (Winn *et al.*, 2011). Además, en algunos casos, se ha recurrido al uso de **Balbes** (Long *et al.*, 2008), programa que automatiza la búsqueda del mejor modelo inicial. Por último **Chainsaw** (Stein, 2008) (también perteneciente a **CCP4**), se ha utilizado para ajustar los modelos iniciales a la secuencia de la proteína a estudiar.

3.1.4.3. Mejora de las fases

Una vez superado el problema de la fase, se puede calcular la función de densidad electrónica. La representación gráfica de esta función origina el mapa de densidad electrónica, el cual proporciona la información necesaria para la construcción del modelo de la proteína. Normalmente, las fases iniciales son muy aproximadas y hay que mejorar el mapa de densidad electrónica que puede no ser predictivo.

En el presente trabajo se han aplicado dos de las técnicas más usadas para la mejora de las fases: aplanamiento de solvente (se genera una máscara que separa proteína y solvente, fijando un valor constante para la densidad del solvente) y promediado de densidad utilizando simetría no cristalográfica (cuando existen varias copias de la proteína dentro de la unidad asimétrica).

3.1.4.4. Construcción del modelo estructural y refinado

Una vez calculado el mejor mapa de densidad ($2F_o - F_c$) posible, se traza la cadena polipeptídica sobre los máximos de densidad, lo cual se suele realizar

de forma manual, aunque en caso de disponer de datos a alta resolución funcionan bien los procedimientos automáticos. En el caso de haber utilizado la técnica de MR, el modelo inicial contiene un sesgo importante hacia la proteína homóloga utilizada, y es necesario alternar una serie de ciclos de refinamiento matemático, con reconstrucción manual del modelo sobre los mapas de densidad electrónica. La reconstrucción manual es la única forma de eliminar errores serios en el modelo que el refinamiento matemático no puede corregir.

El refinamiento implica el ajuste de varios parámetros –**coordenadas de cada átomo** (x, y, z), **factor térmico B** (vibración térmica) y **factor de ocupación** (fracción de moléculas que ocupan una posición)– con el fin de ajustar lo más posible el modelo a la información obtenida de los datos de difracción. Durante el refinado se imponen restricciones estereoquímicas para mantener distancias, ángulos de enlace y torsiones apropiadas.

El refinado matemático normalmente se inicia a través de un modelo de **cuerpo rígido** en el cual la proteína se considera como un todo, y se sigue con modelos estadísticos de **máxima verosimilitud**. En las últimas etapas, se continúa con la adición de ligandos, iones, otras moléculas presentes en el cristal y las moléculas de solvente. Una técnica complementaria es aplicar modelos de **enfriamiento simulado** mediante dinámica molecular, sobre todo en las etapas iniciales ya que tienen un mejor grado de convergencia y puede ser eficaz para eliminar errores importantes en el modelo.

Para evaluar el ajuste del modelo refinado a los datos experimentales, se define el factor de acuerdo R (Ecuación 3.8), que mide la diferencia entre las amplitudes observadas $|F_o|$ y calculadas $|F_c|$ a partir de este modelo.

$$R = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|} \quad \text{(Ecuación 3.8)}$$

Para evitar sesgos derivados de un sobreajuste matemático, se aplica un método estadístico de validación cruzada. Para ello, los datos de difracción se dividen en dos grupos y se introducen dos factores de acuerdo. El primer grupo de reflexiones contiene el 90-95 % de los datos (*working set*, R_{work}) y se utilizan en el refinamiento del modelo. El segundo grupo contiene el 5-10 % restantes (*test set*, R_{free}) y se utiliza para comprobar que el curso del refinamiento realmente conduce a un modelo que se ajusta mejor a los datos experimentales (Brunger, 1992).

Finalmente, en la construcción manual del modelo se utilizan los mapas de densidad electrónica diferencia ($F_o - F_c$). En ellos, los picos de densidad positiva revelan que el modelo está incompleto, mientras que los negativos muestran posiciones incorrectas del mismo.

Se puede dar por válido el modelo cuando no existen restos de densidad electrónica sin modelar, la geometría es correcta y los valores R_{work} y R_{free} son adecuados ($R_{work} < 20\text{-}25\%$ y $R_{free} < 30\%$).

En el transcurso de este trabajo se han empleado los programas **Coot** (Emsley & Cowtan, 2004), para la construcción de los modelos y **Refmac** (Murshudov *et al.*, 2011), para realizar el refinamiento de las estructuras, dentro del paquete de programas **CCP4** (Winn *et al.*, 2011).

3.1.4.5. Validación del modelo

La validación supone la última etapa de la resolución estructural. En ella, debe comprobarse la geometría de todas las moléculas así como su correspondencia con la densidad electrónica. De esta manera, se analiza la posible existencia de choques entre átomos en el cristal, distancias y ángulos de enlace de la proteína, conformación de los rotámeros aceptables para los distintos aminoácidos y torsión del enlace peptídico (ángulos ϕ , φ : N-Ca y ψ , ψ : Ca-C) que debe cumplir los requisitos del diagrama de Ramachandran (Ramachandran *et al.*, 1963) (Figura 3.15).

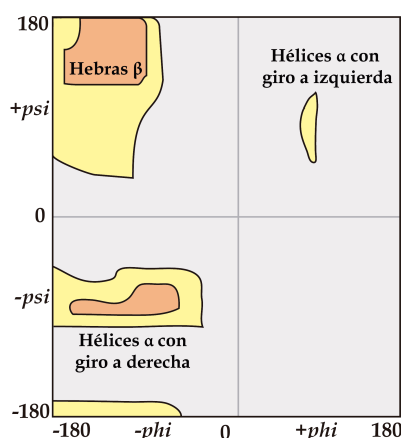


Figura 3.15. Diagrama de Ramachandran. Se destacan las conformaciones más favorables en términos ϕ (φ) y ψ (ψ) del enlace peptídico.

El modelo final se deposita en la base de datos PDB, en donde es sometido a nuevas comprobaciones. Este fichero tiene una extensión *.pdb* y contiene la información cristalográfica y las coordenadas de la macromolécula, que puede ser visualizada con numerosos programas gráficos. Una vez que el fichero es revisado y aprobado, se le asigna un código que lo identifica y pasa a ser de acceso público^{ff}.

En el presente trabajo, la validación se ha realizado mediante los programas **Coot** (Emsley & Cowtan, 2004), **Procheck** (Laskowski *et al.*, 1993), **MolProbity** (Chen *et al.*, 2010) y el visualizador y generador de figuras ha sido **PyMOL** (DeLano, 2002).

^{ff} <http://www.rcsb.org>.

3.2. Otras técnicas físico-químicas

3.2.1. Calorimetría Isotérmica de Titulación

La calorimetría isotérmica de titulación (ITC) se utiliza para evaluar de forma cuantitativa las propiedades termodinámicas de la interacción entre dos componentes (proteína-sustrato, proteína-proteína, etc.). Así, en un solo experimento, se puede determinar la constante de unión (K), la estequiometría de la interacción (n) y la entalpía de unión (ΔH). Además, se pueden determinar indirectamente la entropía (ΔS) y energía libre de Gibbs (ΔG) (ecuaciones 3.9 y 3.10) (Pierce *et al.*, 1999).

$$\Delta G = -RT \ln K_D \quad \text{(Ecuación 3.9)}$$

$$\Delta G = \Delta H - T\Delta S \quad \text{(Ecuación 3.10)}$$

Un calorímetro consta de dos celdas idénticas (celdas de referencia y de muestra) (Figura 3.16). El experimento consiste en medir el calor liberado o absorbido cuando se inyecta uno de los componentes (en nuestro caso la solución de ligando) en la célula muestra que contiene la solución de proteína. Al inyectar, se genera una diferencia de calor entre ambas células la cual debe ser compensada para llegar al equilibrio. Ese calor se evalúa y representa gráficamente, siendo la integral de los picos el calor liberado en cada inyección. Para descontar del calor generado por los efectos de dilución de los componentes, debe realizarse una valoración similar con los mismos parámetros pero sin la proteína.

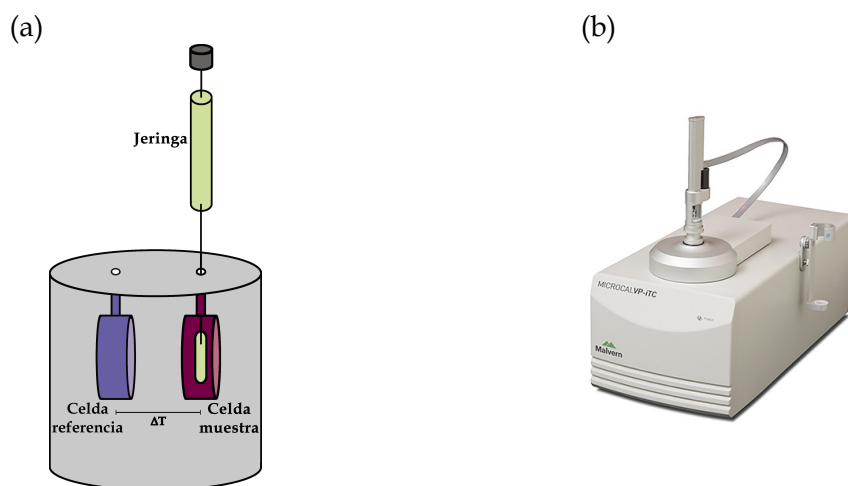


Figura 3.16. (a) Esquema de las celdas que forman un calorímetro. (b) Microcalorímetro usado en el trabajo.

El calorímetro usado en el trabajo ha sido un *MicroCal VP-ITC* (GE Healthcare) (Figura 3.16b).

3.2.2. Ultracentrifugación analítica

La ultracentrifugación analítica (UC) permite determinar el tamaño, la forma global aproximada y el grado de homogeneidad de proteínas u otras macromoléculas en solución (Howlett *et al.*, 2006). Esta técnica se basa en someter una disolución de proteína a un campo centrífugo fuerte, lo que provoca la sedimentación de sus partículas. Además, existen corrientes de difusión que provocan movimientos de las moléculas en sentido contrario, por lo que se genera un gradiente de concentración de partículas. Existen dos tipos de experimentos básicos en UC: **velocidad** y **equilibrio** de sedimentación. Por un lado, en la UC de velocidad de sedimentación, se somete la muestra a un campo centrífugo fuerte tal que la fuerza centrífuga sea mayor que la difusión, por lo que existe un transporte neto hacia el fondo de la celda. Esta técnica permite estimar los coeficientes de sedimentación y difusión, estando estos relacionados con el tamaño y la forma proteica. Por otro lado, en la UC de equilibrio de sedimentación, las muestras se someten a campos centrífugos moderados hasta que se igualan las fuerzas opuestas de centrifugación y difusión. En estas condiciones, las moléculas se distribuyen formando un gradiente de concentración que no varía con el tiempo, el cual no depende de la forma de la proteína pero sí de su masa. Esta técnica permite determinar la masa promedio de las especies que sedimentan y la detección y el análisis cuantitativo de las asociaciones que dan lugar a oligómeros o complejos macromoleculares en solución.

El equipo de UC usado en este trabajo ha sido la ultracentrífuga analítica *Optima XL-A* (Beckman Coulter).

3.3. Otras herramientas bioinformáticas

- **Bioteología-Bioservices Center** (Universidad de Connecticut): Cálculo de pesos moleculares promedio para su uso en UC.
- **BLAST** (*Basic Local Alignment Search Tool*): Alineamiento de secuencias (Altschul *et al.*, 1990).
- **CONCAT32** (*GE Healthcare*): Programa usado para concatenar dos titulaciones consecutivas de ITC llevadas a cabo con igual disolución proteica.
- **DALI** (*Distance ALignment matrix*): Alineamiento estructural de una proteína con otras previamente resueltas y depositadas en el PDB (Holm & Rosenstrom, 2010).
- **eLBOW** (*electronic Ligand Builder and Optimization Workbench*): Mejora de la geometría de ligandos (Moriarty *et al.*, 2009). Está dentro del programa Phenix (Adams *et al.*, 2010).
- **GLYCAM**: Herramienta de modelado de carbohidratos (Kirschner *et al.*, 2008).
- **I-TASSER** (*Iterative Threading ASSEmbly Refinement*): Programa de predicción estructural y funcional (Yang *et al.*, 2015).

- **Origin** software: Perteneciente al equipo de ITC MicroCal.
- **Phenix** (*Python-based Hierarchical ENvironment for Integrated Xtallography*): Automatiza el análisis de datos en cristalografía, desde el análisis de la calidad de los datos, hasta la validación del modelo.
- **PISA**: Análisis de las interfases en oligómeros (Krissinel & Henrick, 2007).
- **Prodrgr**: Genera automáticamente información topológica de pequeñas moléculas como sustratos o ligandos, ajustando la geometría y minimizando la energía. Transforma fichero de coordenadas a distintos formatos. (Schüttelkopf & van Aalten, 2004).
- **ProtParam** (*Protein Identification and Analysis Tools*): Herramienta que permite el cálculo de diferentes parámetros físicos y químicos de una determinada proteína (Wilkins *et al.*, 1999).
- **SEDFIT**: Cálculo de los coeficientes de sedimentación diferencial en UC.
- **SEDNTERP**: Cálculo de los volúmenes específicos parciales a partir de las secuencias de aminoácidos.
- **SUPERPOSE**: Alineamiento estructural basado en la estructura secundaria. Incluido en el paquete de programas **CCP4** (Krissinel & Henrick, 2004).
- **SWEET II**: Modelado y construcción de carbohidratos (www.glycosciences.de).

4. TRABAJOS DE INVESTIGACIÓN

“El científico no busca un resultado inmediato. No espera que sus ideas avanzadas sean fácilmente aceptadas. Su deber es sentar las bases para los que vendrán, señalar el camino”

Nikola Tesla (Croacia, 1856-EEUU 1943)

Inventor, ingeniero mecánico, ingeniero electricista y físico

4.1. Artículo 1:



Crystallization and preliminary X-ray diffraction analysis of the N-terminal domain of *Paenibacillus barcinonensis* xylanase 10C containing the CBM22-1–CBM22-2 tandem

María Ángela Sainz-Polo, Beatriz González, F. I. Javier Pastor and Julia Sanz-Aparicio

Acta Crystallographica Section F (2015). F71, 136–140

4.1.1. Introducción a la cristalización del dominio N-terminal, CBM22-1-CBM22-2, de la xilanasas 10C de *Paenibacillus barcinonensis*

Las hemicelulasas microbianas, principalmente xilanasas, tienen aplicaciones esenciales en la industria debido a su capacidad para modificar y transformar los materiales lignocelulósicos y la pared celular de la biomasa vegetal. Estas enzimas se emplean en una amplia variedad de procesos industriales, desde industria textil y de alimentación hasta biorefinerías. El xilano, principal componente de la hemicelulosa, es un polímero altamente complejo que puede presentar varios tipos de ramificaciones y su degradación enzimática requiere que una gran variedad de enzimas actúen de modo cooperativo para lograr la ruptura completa del mismo. Entre ellas las xilanasas, que son las que rompen el enlace β -1,4 de la cadena principal, desempeñan un papel clave.

Se han caracterizado y clonado xilanasas de varios microorganismos, incluyendo hongos y bacterias. La secuenciación de sus genes ha puesto de manifiesto que muchas de las enzimas descritas suelen tener una estructura modular que consiste en un dominio catalítico unido a uno o más módulos auxiliares con diferentes funciones. El más común de todos estos fragmentos accesorios no catalíticos es el CBM. Este tipo de dominios pueden promover la accesibilidad al xilano. El análisis de la secuencia de ciertas xilanasas de microorganismos termófilos reveló la existencia de fragmentos que originalmente fueron propuestos como dominios termoestabilizantes. Posteriormente fueron catalogados como dominios de unión a xilano y glucanos mixtos del tipo β -1,3-1,4, y se clasificaron dentro de la familia CBM22.

La xilanasas Xyn10C de *P. barcinonensis* es una enzima multimodular de 120 kDa que presenta la arquitectura molecular CBM22/GH10/CBM9. Su dominio N-terminal contiene dos repeticiones de la familia CBM22, mientras que el dominio C-terminal presenta dos módulos homólogos a la familia CBM9. Xyn10C mostró un perfil de actividad similar a los de las enzimas de organismos mesófilos.

La resolución estructural por cristalografía de rayos X del dominio N-terminal conteniendo el tándem CBM22-1-CBM22-2 contribuye a esclarecer el papel de los dominios auxiliares no catalíticos en este tipo de enzimas microbianas y supone la única estructura de un tándem de CBM22 resuelta hasta el momento. El conocimiento de la misma ayudará a comprender cual es el papel de la multiplicidad de CBM en la función biológica de diferentes xilanasas microbianas.

Los objetivos de este trabajo fueron los que se describen a continuación:

1. Definir un protocolo de expresión y purificación de una construcción que contenga la región CBM22-1-CBM22-2 y permita obtener muestra de calidad suficiente para poder llevar a cabo los posteriores experimentos de cristalización.
2. Establecer las condiciones óptimas de cristalización y difracción de los cristales.
3. Interpretación de los datos estadísticos de difracción para garantizar la posterior resolución estructural del dominio CBM22-1-CBM22-2.

La autora de esta tesis ha realizado:

- a. Clonaje, expresión y purificación del dominio N-terminal de Xyn10C conteniendo CBM22-1-CBM22-2.
- b. Cristalización del dominio N-terminal.
- c. Toma de datos de difracción en el sincrotrón ESRF de Grenoble, en la línea ID29.
- d. Procesado de los datos y valoración de su calidad para la posterior resolución estructural.

4.1.2. Publicación

research communications



STRUCTURAL BIOLOGY
COMMUNICATIONS

ISSN 2053-230X

Crystallization and preliminary X-ray diffraction analysis of the N-terminal domain of *Paenibacillus barcinonensis* xylanase 10C containing the CBM22-1–CBM22-2 tandem

María Ángela Sainz-Polo,^a Beatriz González,^a F. I. Javier Pastor^b and Julia Sanz-Aparicio^{a*}

Received 25 November 2014
Accepted 17 December 2014

Keywords: carbohydrate-binding domain; CBM22; *Paenibacillus*; bacterial xylanase; xylan-binding domain.

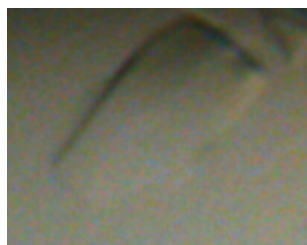
^aDepartment of Crystallography and Structural Biology, Institute of Physical Chemistry Rocasolano, CSIC, Serrano 119, 28006 Madrid, Spain, and ^bDepartment of Microbiology, Faculty of Biology, University of Barcelona, Avenida Diagonal 643, 08028 Barcelona, Spain. *Correspondence e-mail: xsainz@iqfr.csic.es

A construct containing the CBM22-1–CBM22-2 tandem forming the N-terminal domain of *Paenibacillus barcinonensis* xylanase 10C (Xyn10C) has been purified and crystallized. A xylan-binding function and an affinity for mixed β -1,3/ β -1,4 glucans have previously been demonstrated for some members of the CBM22 family. The sequence of the tandem is homologous to the N-terminal domains found in several thermophilic enzymes. Crystals of this tandem were grown by the streak-seeding method after a long optimization strategy. The structure has been determined by molecular replacement to a resolution of 2.43 Å and refinement is under way. This study represents the first structure containing two contiguous CBM22 modules, which will contribute to a better understanding of the role that this multiplicity plays in fine-tuning substrate affinity.

1. Introduction

Microbial hemicellulases, primarily xylanases (EC 3.2.1.8), have essential applications in industry owing to their ability to modify and transform lignocellulose and cell-wall materials of vegetal biomass. These enzymes are used in a wide variety of industrial processes such as the textile, baking and food industries to improve the properties of the products obtained. In the pulp-bleaching process, the application of xylanases reduces the chlorine requirement and thus notably contributes to the sustainability of the pulp and paper industry (Verma & Satyanarayana, 2012). Xylan, the major component of cell walls, is a highly complex polymer and therefore its enzymatic degradation requires a large variety of cooperatively acting enzymes to achieve an efficient process (Biely, 1985; de Vries & Visser, 2001). Among these enzymes, xylanases, which cleave the internal β -1,4 bonds within the main chain of xylan, play a key role.

Xylanases from several microorganisms, including fungi and bacteria, have been characterized and cloned. Sequencing of these genes has revealed that many of the encoded enzymes have a modular structure consisting of a catalytic domain attached to one or more ancillary domains with different functions (Collins *et al.*, 2005). The most common noncatalytic domains in modular xylanases are carbohydrate-binding modules (CBMs; Boraston *et al.*, 2004) that may promote accessibility to xylan, which occurs in spatial proximity to cellulose in natural substrates (Hervé *et al.*, 2010; Gilbert *et al.*, 2013). Analysis of the sequences of xylanases from thermophilic microorganisms identified other conserved sequences that were originally proposed to act as thermostabilizing



© 2015 International Union of Crystallography

136 doi:10.1107/S2053230X14027496

Acta Cryst. (2015). F71, 136–140

Table 1
Macromolecule-production information.

| | |
|--|--|
| Source organism | <i>P. barcinonensis</i> |
| DNA source | Genomic DNA/pET-11a |
| Forward primer† | CGCGGATCCGCGAAGGGGGGATATTC |
| Reverse primer† | GGCTCGAGTTAAGCAATCGCTTCGGTTGC |
| Cloning vector | pGEX-4T-2 |
| Expression vector | pGEX-4T-2 |
| Expression host | <i>E. coli</i> BL21 (DE3) |
| Complete amino-acid sequence of the construct produced | MASAAKAGDILLSHSFEEGTTQGTARGVKVDV-TAEQAYQKQSLQTTGRTEAWNGPSLSLTDVV-HKNEVVEISGYVKLVAGSAPADLKFTVERRDG-NGDTQYDVNAEQVTDQKWKQLQGQSYEQG-SSLLLYLESTDAKAAAYLLDEFQIRLVKAAPEN-PGEPGEAGQALFKADFEDGNIGNWRARGTEKL-EVVSIGIGHNSNRSLKTSRSEYTHGPLEVLP-YLQKGSTVHISFWAMYDEGPATQVINGSLEKE-FNRDTANLEYAMFASTTLNKGQWKIEADIIV-PAESTGISGLRMYAETPWKQSEVTETDTIPF-VYDDVQITATEAIAIE |

† BamHI and XhoI restriction sites are underlined.

domains (TSDs; Sunna *et al.*, 2000; Blanco *et al.*, 1999) but were later shown to act as xylan-binding domains which bind xylan and mixed β -1,3/ β -1,4-glucan (Meissner *et al.*, 2000) and were classified into family 22 of carbohydrate-binding modules (CBM22).

Xylanase 10C (Xyn10C) from *Paenibacillus barcinonensis* (Sánchez *et al.*, 2005) is a 120 kDa multimodular enzyme that presents the CBM22/GH10/CBM9 modular architecture found in a subset of large GH10 xylanases (Collins *et al.*, 2005). Its N-terminal domain contains two repeats of family 22 carbohydrate-binding modules, both presenting the motifs previously identified as essential determinants for xylan binding (Xie *et al.*, 2001). In addition to its GH10 catalytic domain, two domains homologous to family 9 of cellulose-binding domains (CBM9) are located in the C-terminal region. Xyn10C showed an activity profile similar to those of enzymes from mesophilic microorganisms (Blanco *et al.*, 1999). Here, we report the purification, crystallization and preliminary X-ray crystallographic analysis of the N-terminal domain of *P. barcinonensis* Xyn10C containing the CBM22-1–CBM22-2 tandem (Xyn10C-Nterm). This represents the first structure of two contiguous CBM22 domains to be solved. Elucidation of this structure will contribute to understanding the role of multiplicity in the biological function of different microbial xylanases.

2. Materials and methods

2.1. Macromolecule production

The cDNA of Xyn10C-Nterm was amplified from a XYN10C-Nterm/pET-11a vector (UniProt accession No. O69230) using the forward and reverse oligonucleotides given in Table 1. The Xyn10C-Nterm cDNA generated included restriction sites for BamHI and XhoI enzymes that were used to insert the cDNA into the pGEX-4T-2 vector (GE Healthcare). The sequence of the Xyn10C-Nterm/pGEX-4T-2 plasmid was verified by DNA sequencing.

For protein expression, *Escherichia coli* BL21 (DE3) cells were transformed with the Xyn10C-Nterm/pGEX-4T-2 plasmid and grown in ampicillin-supplemented LB medium ($50 \mu\text{g ml}^{-1}$) at 310 K until the absorbance at 600 nm reached 0.8. Expression of the GST-fused protein was induced with 0.3 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) for 16 h at 289 K. Cells were collected by centrifugation and kept at 253 K until use.

The cells were suspended in 20 mM Tris pH 8.0, 100 mM NaCl (buffer A) and disrupted by sonication. The cell lysate was centrifuged and the resulting supernatant was filtrated. The clarified lysate was mixed with GST-trap resin (GenScript) equilibrated in buffer A and incubated for 1 h at 277 K. The resin was generously washed with buffer A. To cleave the GST-fused protein, the resin was incubated with thrombin (10 units per milligram of protein) overnight at 277 K. The cleaved Xyn10C-Nterm was collected, diluted (to reduce the salt to 50 mM NaCl) and subjected to ion-exchange chromatography (IEX) on a Mono Q column (ÄKTA; GE Healthcare). The protein was eluted using a gradient buffer of 50 mM to 1 M NaCl in 20 mM Tris pH 8.0. The protein concentration was estimated using a molar extinction coefficient (ϵ) of $57\,870 \text{ M}^{-1} \text{ cm}^{-1}$ at 280 nm with a NanoDrop ND-1000 spectrophotometer and was verified by SDS–PAGE (Fig. 1). Purified Xyn10C-Nterm was concentrated to 47 mg ml^{-1} for protein crystallization assays. The purity of the protein was determined by SDS–PAGE (Laemmli, 1970).

2.2. Crystallization

In order to find a suitable protein concentration, the Pre-Crystallization Test (PCT; Hampton Research) was used, which gave a starting point of $30\text{--}50 \text{ mg ml}^{-1}$. Initial crystallization conditions of the purified protein samples were investigated by high-throughput techniques with a NanoDrop

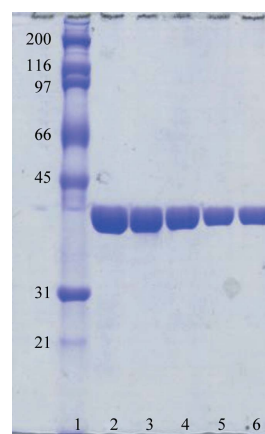


Figure 1
SDS–PAGE analysis of purified Xyn10C-Nterm on 12% polyacrylamide gel. Lane 1, molecular-weight protein markers (labelled in kDa); lanes 2–6, purified Xyn10C-Nterm (Mono Q column fractions). The molecular weight calculated from the sequence is 37.43 kDa.

research communications

robot (Innovadyne Technologies Inc.) using the commercially available Index, SaltRx and Crystal Screen from Hampton Research and The PACT Suite and The JCSG Suite from Qiagen. The assays were carried out by the sitting-drop vapour-diffusion method at 291 K in Innovaplate SD-2 microplates (Innovadyne Technologies Inc.) by mixing 250 nl protein solution with 250 nl precipitant solution and equilibrating against 60 μ l well solution. Optimization experiments were performed as further sitting-drop experiments by mixing 0.5–1 μ l protein solution with 0.5–2 μ l precipitant solution and equilibrating against 500 μ l well solution in Cryschem plates (Hampton Research).

2.3. Data collection and processing

All crystals were soaked in precipitant solution containing an additional 20% (v/v) glycerol (Garman & Mitchell, 1996) before being flash-cooled to 100 K in liquid nitrogen. Crystals were tested using synchrotron radiation. Diffraction data sets were collected on the ID29 beamline at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. The data sets were processed using *iMosflm* (Battye *et al.*,

2011) and *AIMLESS* (Evans, 2006) as distributed in the *CCP4* suite (Winn *et al.*, 2011).

3. Results and discussion

Xyn10C-Nterm was produced from *E. coli* cells in a soluble form and was purified in two steps (GST-tagged protein purification resin and ion-exchange chromatography) with a high degree of purity (Fig. 1, Table 1). A band in the SDS-PAGE analysis showing the expected molecular weight of the affinity-captured protein, which was later cleaved by thrombin, confirmed the identity of the sample. The purified protein was submitted to crystallization trials, with several conditions from commercial screens containing 25–30% PEG 3350 or PEG 8000 and pH 5.5–6.5 resulting in spherulites (Fig. 2a). Fine grid screening of these conditions did not improve the results and only the screening of additives led to the formation of small needles, which grew from the spherulites when using 25% PEG 3350 as precipitant and 2% (v/v) 1,6-hexanediol as an additive (Fig. 2b). Additional experiments were performed using different protein:precipitant: additive ratios and sampling the pH, and always led to the

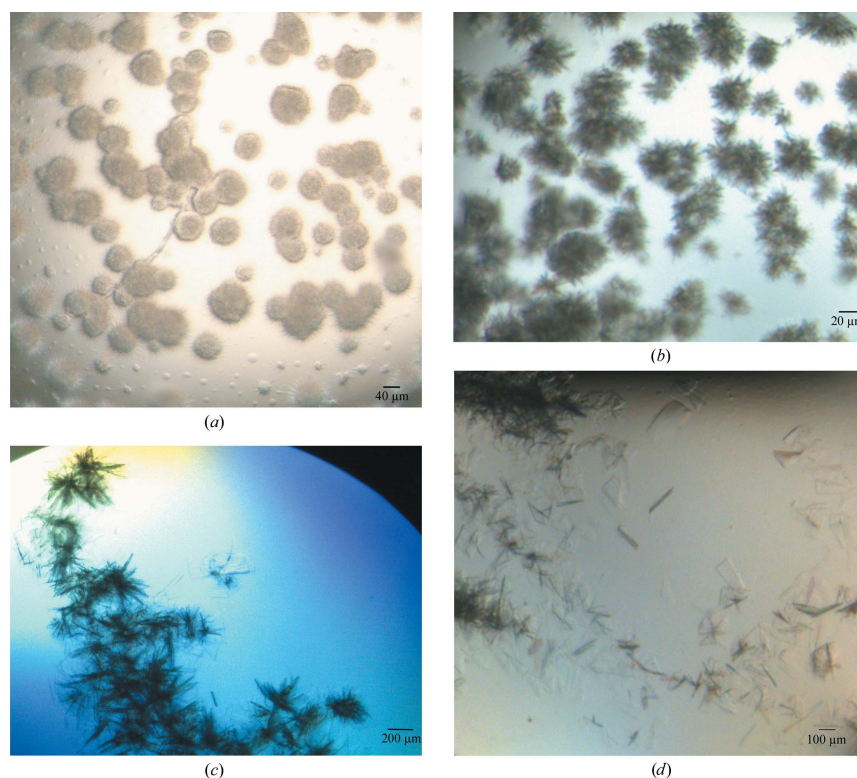


Figure 2
Xyn10C-Nterm crystals obtained from the following conditions. (a) 32 mg ml⁻¹ protein solution and 25% PEG 3350, 0.2 M NaCl, 0.1 M bis-tris pH 5.5; 1:1 ratio. (b) 47 mg ml⁻¹ protein solution and 28% PEG 3350, 0.2 M NaCl, 0.1 M bis-tris pH 5.5; 1:1 ratio; 2% (v/v) 1,6-hexanediol was used as an additive. (c) 47 mg ml⁻¹ protein solution and 25% PEG 3350, 0.2 M NaCl, 0.1 M bis-tris pH 5.5, 2% (v/v) 1,6-hexanediol; 1:1 ratio; streak-seeding from spherulites. (d) 47 mg ml⁻¹ protein solution and 23% PEG 3350, 0.2 M NaCl, 0.1 M bis-tris pH 5.5, 2% (v/v) 1,6-hexanediol; 1:4 ratio.

Table 2
Crystallization.

| | |
|--|--|
| Method | Vapour diffusion (sitting drop) and streak-seeding |
| Plate type | Crychem plates |
| Temperature (K) | 291 |
| Protein concentration (mg ml ⁻¹) | 47 |
| Buffer composition of protein solution | 20 mM Tris pH 8.0, 80 mM NaCl |
| Composition of reservoir solution | 23% PEG 3350, 0.2 M NaCl, 0.1 M bis-tris pH 5.5 |
| Volume (μl) and ratio of drop | 0.5:2:0.3 (protein:reservoir: 2% 1,6-hexanediol) |
| Volume of reservoir (μl) | 500 |

same crystallization pattern. According to Bergfors (2003), this association trend is ideal for carrying out streak-seeding experiments and, indeed, very thin plates grew from streak-seeding (Fig. 2c). The best crystals were obtained by mixing 47 mg ml⁻¹ protein solution with 23% PEG 3350, 0.2 M NaCl, 0.1 M bis-tris pH 5.5, 2% (v/v) 1,6-hexanediol at 291 K, leaving the mixture to equilibrate overnight and then streak-seeding from a drop containing spherulites with a whisker probe. Isolated thin plates appeared after 2–3 months (Fig. 2d). Crystallization information is summarized in Table 2.

The optimized crystals were analysed using a synchrotron-radiation source to obtain high-resolution data (Fig. 3). The best Xyn10C-Nterm crystal diffracted to a resolution of 2.43 Å. The crystals belonged to space group *P*₂₁, with unit-cell parameters *a* = 84.170, *b* = 110.359, *c* = 118.522 Å, *β* = 90.63° (Table 3). As observed by SDS–PAGE and calculated from sequence analysis, the molecular weight of the monomer is 37 kDa (Fig. 1). Assuming a reasonable Matthews coefficient value within the range 3.68–1.84 Å³ Da⁻¹ (Matthews, 1968), corresponding to 66–33% solvent content,

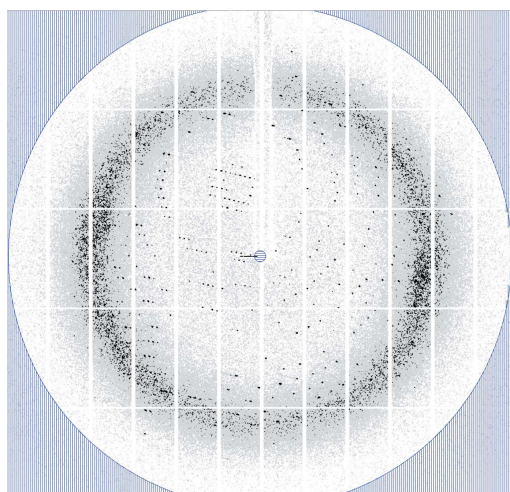


Figure 3
X-ray diffraction pattern of Xyn10C-Nterm obtained using a synchrotron source. The maximum observed resolution is 2.43 Å.

Table 3
Data collection and processing.

| | |
|--|------------------------------------|
| Values in parentheses are for the outer shell. | |
| Diffraction source | Beamline ID29, ESRF |
| Wavelength (Å) | 0.979235 |
| Temperature (K) | 120 |
| Detector | PILATUS 6M-F |
| Crystal-to-detector distance (mm) | 499.68 |
| Rotation range per image (°) | 0.1 |
| Total rotation range (°) | 360 |
| Exposure time per image (s) | 0.037 |
| Space group | <i>P</i> ₂ ₁ |
| <i>a</i> , <i>b</i> , <i>c</i> (Å) | 84.170, 110.359, 118.522 |
| <i>β</i> (°) | 90.63 |
| Mosaicity (°) | 0.44 |
| Resolution range (Å) | 84.16–2.43 (2.48–2.43) |
| Total No. of reflections | 510618 (29165) |
| No. of unique reflections | 81116 (4468) |
| Completeness (%) | 99.4 (99.3) |
| Multiplicity | 6.3 (6.5) |
| <i>I</i> (σ(<i>I</i>)) | 8.6 (2.2) |
| <i>R</i> _{int} | 0.060 (0.397) |
| Overall <i>B</i> factor from Wilson plot (Å ²) | 37.24 |

† $R_{p.i.m.} = \sum_{hkl} (1/[N(hkl) - 1])^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $\langle I(hkl) \rangle$ is the average of symmetry-related observations of a unique reflection.

the presence of 4–8 molecules in the asymmetric unit should be expected. We investigated the local symmetry relating the subunits in the asymmetric unit using *POLARRFN* (Kabsch, 1976) from the *CCP4* package. Several self-rotation functions were computed in the resolution range 15–3.5 Å with Patterson vectors of 25–35 Å radius of integration. The stereographic projection of the $\kappa = 180^\circ$ section (Fig. 4) revealed the presence of noncrystallographic twofold symmetry almost parallel to the *a* and *c* axes.

Structure determination is in progress using *Phaser* (McCoy *et al.*, 2007) for molecular replacement. The templates used were CBM22-1 and CBM22-2 from *Clostridium thermocellum*

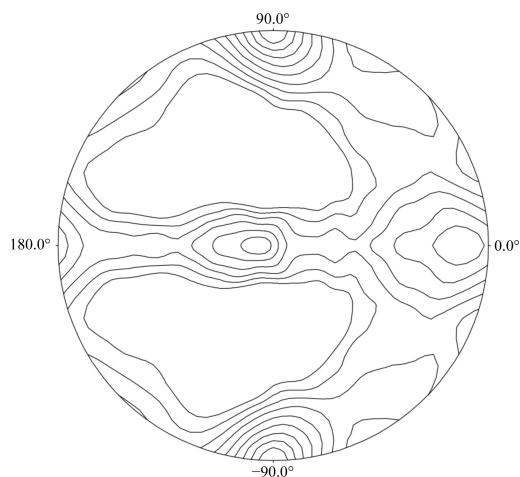


Figure 4
Plot of the self-rotation function of Xyn10C-Nterm crystals in the $\kappa = 180^\circ$ section.

research communications

xylanase Y (PDB entries 2w5f and 1dyo; Najmudin *et al.*, 2010; Charnock *et al.*, 2000), which show about 20% sequence identity with respect to each Xyn10C-Nterm subdomain. A solution with six molecules within the asymmetric unit was obtained and preliminary structural refinement with *REFMAC* (Murshudov *et al.*, 2011) decreased the *R* factor to 0.37 ($R_{\text{free}} = 0.42$). Model building and further refinement are ongoing. All data-collection statistics are shown in Table 3.

Acknowledgements

This work was supported by grants BIO2010-20508-C04-03 and BIO2013-48779-C4-2-R from Dirección General de Investigación, MICINN. This is a product of the Project 'Factoría Española de Cristalización' Ingenio/Consolider 2010. MASP is supported by a JAE-PreDoc fellowship from CSIC. We also thank the ESRF for beamtime and the ID29 staff for providing assistance with data collection.

References

- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Bergfors, T. (2003). *J. Struct. Biol.* **142**, 66–76.
- Biely, P. (1985). *Trends Biotechnol.* **3**, 286–290.
- Blanco, A., Diaz, P., Zueco, J., Parascandola, P. & Pastor, F. I. J. (1999). *Microbiology*, **145**, 2163–2170.
- Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). *Biochem. J.* **382**, 769–781.
- Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M. A., Davies, G. J. & Fontes, C. M. G. A. (2000). *Biochemistry*, **39**, 5013–5021.
- Collins, T., Gerday, C. & Feller, G. (2005). *FEMS Microbiol. Rev.* **29**, 3–23.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Garman, E. F. & Mitchell, E. P. (1996). *J. Appl. Cryst.* **29**, 584–587.
- Gilbert, H. J., Knox, J. P. & Boraston, A. B. (2013). *Curr. Opin. Struct. Biol.* **23**, 669–677.
- Hervé, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J. & Knox, J. P. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 15293–15298.
- Kabsch, W. (1976). *Acta Cryst.* **A32**, 922–923.
- Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Meissner, K., Wassenberg, D. & Liebl, W. (2000). *Mol. Microbiol.* **36**, 898–912.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Najmudin, S., Pinheiro, B. A., Prates, J. A. M., Gilbert, H. J., Romão, M. J. & Fontes, C. M. G. A. (2010). *J. Struct. Biol.* **172**, 353–362.
- Sánchez, M. M., Fritze, D., Blanco, A., Spröer, C., Tindall, B. J., Schumann, P., Kroppenstedt, R. M., Diaz, P. & Pastor, F. I. J. (2005). *Int. J. Syst. Evol. Microbiol.* **55**, 935–939.
- Sunna, A., Gibbs, M. D. & Bergquist, P. L. (2000). *Biochem. J.* **346**, 583–586.
- Verma, D. & Satyanarayana, T. (2012). *Bioresour. Technol.* **117**, 360–367.
- Vries, R. P. de & Visser, J. (2001). *Microbiol. Mol. Biol. Rev.* **65**, 497–522.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Xie, H., Bolam, D. N., Nagy, T., Szabó, L., Cooper, A., Simpson, P. J., Lakey, J. H., Williamson, M. P. & Gilbert, H. J. (2001). *Biochemistry*, **40**, 5700–5707.

4.1.3. Resultados y discusión

El dominio N-terminal de Xyn10C (CBM22-1-CBM22-2) se expresó a partir de células de *Escherichia coli* de forma soluble y se purificó en dos pasos (cromatografía de afinidad mediante fusión a cola de GST[®] en resina y cromatografía de intercambio iónico) con un alto grado de pureza.

Un rastreo inicial de condiciones de cristalización se llevó a cabo con una extensa búsqueda de más de 480 soluciones comerciales con la ayuda del robot *NanoDrop (Innovadine Technologies Inc.)* disponible en nuestro laboratorio. Varias de estas condiciones (25-30 % (m/v) PEG 3350 o PEG 8000 y pH 5,5-6,5) generaron esferulitas. El intento de optimización de estos cristales mediante variación de concentración de precipitante o pH, no condujo a una mejora de la calidad obteniéndose sólo esferulitas. Se realizaron por tanto experimentos con 96 aditivos diferentes, observándose la aparición de unas micro agujas cristalinas a partir de las esferulitas, cuando se utilizó 1,6-hexanodiol como aditivo. Este tipo de hábito cristalino es ideal para llevar a cabo experimentos de “sembrado por arrastre”, que teóricamente permiten transferir micro semillas cristalinas a una solución que contenga proteína fresca equilibrada con el mismo agente precipitante. Plaquititas muy finas, pero de buena calidad, crecieron después de dos meses de realizada la siembra. Las condiciones de cristalización finales fueron: 47 mg ml⁻¹ de proteína con 23 % PEG 3350 (m/v), 0,2 M NaCl, 0,1 M bis-tris pH 5,5, 2 % (v/v) 1,6-hexanodiol a 291 K, con una proporción de proteína/precipitante/aditivo: 0,5:2:0,3.

Los mejores cristales fueron analizados utilizando una fuente de radiación sincrotrón para obtener datos a alta resolución. El modo de crioprotección de los mismos fue sumergirlos en las condiciones de cristalización más un 20 % (m/v) de glicerol para una posterior criogenización en nitrógeno líquido a 100 K. Varios conjuntos de datos se recogieron a 100 K en el sincrotrón ESRF de Grenoble, en la línea ID29. Los cristales pertenecen al grupo espacial $P2_1$, con parámetros de celdilla $a = 84 \text{ \AA}$, $b = 110 \text{ \AA}$, $c = 118 \text{ \AA}$, $\beta = 90,63^\circ$ y difractaron a una resolución máxima de $2,43 \text{ \AA}$. El contenido de solvente es de 33-66 %, esperándose, por tanto, la presencia de 4-8 moléculas dentro en la unidad asimétrica.

[®] GST: Glutación-S-transferasa.

4.2. Artículo 2:



Exploring multimodularity in plant cell wall deconstruction: Structural and functional analysis of Xyn10C containing the CBM22-1-CBM22-2 tandem

María Ángela Sainz-Polo, Beatriz González, Margarita Menéndez, F. I. Javier Pastor and Julia Sanz-Aparicio

The Journal of Biological Chemistry (2015) **290**, 28: 17116-17130

4.2.1. Introducción al análisis estructural y funcional del dominio N-terminal de la xilanasa Xyn10C de *Paenibacillus barcinonensis*

El complejo entramado de polímeros que forma la pared celular vegetal hace muy complicada su degradación. Para afrontar este problema, la naturaleza ha creado una gran variedad de enzimas de diferentes familias que pueden actuar de un modo coordinado para degradar la biomasa. Además, algunas de estas enzimas han desarrollado una composición modular con dominios adicionales al catalítico que actúan aumentando la accesibilidad del dominio catalítico al sustrato. La mayor parte de estos dominios auxiliares son CBM. Su presencia en una amplia gama de GH muestra claramente su importancia funcional, permitiendo la interacción de sustratos poco accesibles o insolubles con el sitio activo del dominio catalítico. Sin embargo, su modo exacto de acción es aún desconocido.

Algunas enzimas contienen múltiples copias de CBM que muestran la misma especificidad perteneciendo incluso a la misma familia. En esas proteínas multimodulares, la presencia de varias copias de CBM homólogos sugiere un sutil mecanismo de regulación de la especificidad enzimática. Así, para comprender el mecanismo mediante el cual las enzimas modulares interactúan con su sustrato y para determinar la relación funcional entre la unión de los carbohidratos y la catálisis, se requiere conocer en detalle los determinantes estructurales que presentan estos CBM.

La xilanasa Xyn10C pertenece al secretoma de *P. barcinonensis*, un potente microorganismo xilanolítico que produce un grupo variado de enzimas, incluyendo xilanasas de la familia GH10, GH11 y GH30, las cuales se han clonado y caracterizado. Xyn10C, caracterizada previamente, mostró una estructura modular CBM22-1-CBM22-2-GH10-CBM9a-CBM9b.

Actualmente, diversos estudios sugieren que la presencia de sistemas tipo tándem en GH podría revelar un mecanismo mucho más sofisticado de lo que se creía, descartando así la idea clásica de que una disposición homogénea de CBM en una proteína suponga, meramente, la existencia de efectos de multivalencia y/o avidéz frente a una estructura polimérica heterogénea. De este modo, el tándem de módulos CBM22 presente en el dominio N-terminal de Xyn10C es un sistema atractivo de estudio para poder comprender mejor los mecanismos de especificidad regulados por CBM en enzimas multimodulares.

El dominio N-terminal de Xyn10C fue expresado, purificado y cristalizado previamente. Los objetivos del presente trabajo se han centrado en continuar con el estudio de dichos dominios no catalíticos, investigar sus capacidades de unión y sugerir un posible mecanismo de regulación de especificidad dentro del contexto de Xyn10C. Los objetivos concretos se recogen a continuación:

1. Resolución estructural del tándem CBM22-1-CBM22-2 (Xyn10C-XBD) e interpretación de los resultados.
2. Definir un protocolo de expresión y purificación de los módulos CBM22-1 y CBM22-2 aislados, de la proteína completa Xyn10C así como del dominio catalítico GH10 (Xyn10C-CD).
3. Realizar ensayos de unión con sustratos específicos mediante ITC y a través del análisis de geles de afinidad de cada proteína purificada.
4. Establecer las condiciones óptimas de cristalización y difracción de Xyn10C-XBD con sustratos específicos.
5. Establecer las condiciones óptimas de cristalización y difracción de CBM22-1 y CBM22-2 con y sin sustratos afines.
6. Resolución estructural de CBM22-1 y CBM22-2 y sus complejos. Interpretación de los datos.
7. Establecer las condiciones óptimas de cristalización y difracción del dominio catalítico, Xyn10C-CD.
8. Resolución estructural de Xyn10C-CD e interpretación de los datos.
9. Interpretación de los datos estructurales y bioquímicos obtenidos de Xyn10C y análisis de la funcionalidad de los distintos dominios en la actividad de Xyn10C. En base a ello, análisis del papel de la multimodularidad en el mecanismo de degradación de la pared celular vegetal.

La autora de esta tesis ha realizado:

- a. Resolución estructural de Xyn10C-XBD.
- b. Purificación de Xyn10C y Xyn10C-CD y clonaje, expresión y purificación de CBM22-1 y CBM22-2.
- c. Cristalización de Xyn10C-CD, CBM22-1, CBM22-2, y de sus complejos: CBM22-2/Xyl4, CBM22-2/Xyl3 y CBM22-2/Glc-4Glc-4Glc-3Glc.
- d. Toma de datos de difracción en varias estaciones de radiación sincrotrón europeas de las muestras cristalizadas en el apartado “c” y resolución estructural de las mismas.
- e. Interpretación de los resultados obtenidos y del papel de la multimodularidad en Xyn10C.

4.2.2. Publicación

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 290, NO. 28, PP. 17116–17130, JULY 10, 2015
© 2015 BY THE AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY, INC. PUBLISHED IN THE U.S.A.

Exploring Multimodularity in Plant Cell Wall Deconstruction STRUCTURAL AND FUNCTIONAL ANALYSIS OF Xyn10C CONTAINING THE CBM22-1–CBM22-2 TANDEM*

Received for publication, April 15, 2015, and in revised form, May 20, 2015. Published, JBC Papers in Press, May 22, 2015, DOI 10.1074/jbc.M115.659300

M. Angela Sainz-Polo^{‡1}, Beatriz González[‡], Margarita Menéndez[§], F. I. Javier Pastor[§], and Julia Sanz-Aparicio^{‡2}

From the Departamentos de [‡]Cristalografía y Biología Estructural y [§]Química Física Biológica, Instituto de Química-Física Rocasolano, Consejo Superior de Investigaciones Científicas, Serrano 119, 28006-Madrid and the [§]Departamento de Microbiología, Facultad de Biología, Universidad de Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain

Background: Multimodularity is essential in most plant cell wall recycling enzymes.

Results: The CBM22-1–CBM22-2 tandem possesses extraordinary plasticity and two sites with different affinities. CBM22-2 exhibits novel specificity within CBM22s.

Conclusion: Depolymerization of highly substituted arabinoxylans by Xyn10C and a delivery strategy mediated by Xyn10C-XBD are proposed.

Significance: Elucidating the mechanisms regulating specificity through multimodularity is crucial for producing efficient biocatalysts for biomass deconstruction.

Elucidating the molecular mechanisms regulating multimodularity is a challenging task. *Paenibacillus barcinonensis* Xyn10C is a 120-kDa modular enzyme that presents the CBM22/GH10/CBM9 architecture found in a subset of large xylanases. We report here the three-dimensional structure of the Xyn10C N-terminal region, containing the xylan-binding CBM22-1–CBM22-2 tandem (Xyn10C-XBD), which represents the first solved crystal structure of two contiguous CBM22 modules. Xyn10C-XBD is folded into two separate CBM22 modules linked by a flexible segment that endows the tandem with extraordinary plasticity. Each isolated domain has been expressed and crystallized, and their binding abilities have been investigated. Both domains contain the R(W/Y)YYE motif required for xylan binding. However, crystallographic analysis of CBM22-2 complexes shows Trp-308 as an additional binding determinant. The long loop containing Trp-308 creates a platform that possibly contributes to the recognition of precise decorations at subsite S2. CBM22-2 may thus define a subset of xylan-binding CBM22 modules directed to particular regions of the polysaccharide. Affinity electrophoresis reveals that Xyn10C-XBD binds arabinoxylans more tightly, which is more apparent when CBM22-2 is tested against highly substituted xylan. The crystal structure of the catalytic domain, also reported, shows the capacity of the active site to accommodate xylan substitutions at almost all subsites. The structural differences found at both Xyn10C-XBD domains are consistent with the isothermal titration calorimetry experiments showing two

sites with different affinities in the tandem. On the basis of the distinct characteristics of CBM22, a delivery strategy of Xyn10C mediated by Xyn10C-XBD is proposed.

Plant cell walls are highly complicated constructions containing an intricate network of polymers that are recalcitrant to degradation. Moreover, they are very dynamic and change their structures and composition depending on cell development and environmental conditions. To cope with this extraordinary complexity, nature has created a large variety of enzymes that act synergistically to deconstruct biomass (1). An interesting molecular mechanism developed by many plant cell wall hydrolases is their modular composition, with additional domains that increase accessibility of its catalytic partner to the substrate. Most of these supplementary domains are carbohydrate-binding modules (CBMs).³ Their presence in a broad range of hydrolases clearly highlights their functional importance, although their precise mode of action is still unknown. In fact, the occurrence of several copies of homologous CBMs within some enzymes suggests an additional subtle mechanism of regulation in enzymatic specificity. Structural studies from isolated CBMs have revealed that substrate recognition and binding by these domains are extremely specific. To understand the mechanism by which modularity enhances the interaction of enzymes with their substrate and to determine the functional relationship between carbohydrate binding and catalysis, more structural information on multiplicity is required at the atomic level.

CBMs are generally considered as noncatalytic modules that potentiate the activity of carbohydrate-active enzymes by bringing them in close proximity to their substrates. Most

* This work was supported by Grants BIO2010-20508-C04-03, BIO2013-48779-C4-2-R, CTQ2013-48995-C2-2-R, and BFU2012-36825 from the Spanish Ministry of Science and Innovation. The authors declare that they have no conflicts of interest with the contents of this article.

The atomic coordinates and structure factors (codes 4XUP, 4XUO, 4XUN, 4XUR, 4XUQ, 4XUT, and 4W8L) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

¹ Recipient of a Junta de Ampliación de Estudios predoctoral fellowship from the Spanish National Research Council.

² To whom correspondence should be addressed. Tel.: 34-91-561-9400; Fax: 34-91-564-2431; E-mail: xjulia@iqfr.csic.es.

³ The abbreviations used are: CBM, carbohydrate-binding module; a.u., asymmetric unit; r.m.s.d., root mean square deviation; RthXyn10B, *R. thermocellum* xylanase 10B; ITC, isothermal titration calorimetry; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; PDB, Protein Data Bank.

TABLE 1

Oligonucleotides employed

Forward (F) and reverse (R) sequences include the restriction sites in boldface letters.

| Protein | Orientation | Sequence |
|------------|-------------|--|
| Xyn10C | F | 5'-GGAGAAC CATATG CGCAAGCGCAGCG-3' |
| | R | 5'-TGGCG GGATCC GAGCTTGAACA-3' |
| Xyn10C-CD | F | 5'-GATTC CATATG CAAAAAAATATTCGG-3' |
| | R | 5'-TGC GGATCC CTAATTCGATATACG-3' |
| Xyn10C-XBD | F | 5'-CGC GGATCC ATGGCAAGCGCAGCGAAGGCGG-3' |
| | R | 5'-GGC CTCGAG TTATTCATAGCAATCGCTTCGG-3' |
| CBM22-1 | F | 5'-CGC GGATCC ATGGCAAGCGCAGCGAAGGCGG-3' |
| | R | 5'-GGC CTCGAG TTACGCAAGCTTTAACGAGACGAATC-3' |
| CBM22-2 | F | 5'-CGC GGATCC GGTGAGCCGGGAGAGCAGGAC-3' |
| | R | 5'-GGC CTCGAG TTATTCATAGCAATCGCTTCGG-3' |

CBMs are part of modular carbohydrate hydrolases, although several examples of CBMs linked to cellulosomal scaffolding proteins or not linked to a hydrolase at all, *i.e.* the so-called "orphan CBMs," have been described (2). Based on amino acid sequence similarity, they are classified into different families (CAZy); currently 71 are known, and the number is still increasing (3). CBMs show substantial differences in their ligand specificity and binding properties (4, 5). They recognize and bind a diversity of plant carbohydrates, including crystalline and non-crystalline cellulose, soluble and insoluble xylan, chitin, β -glucans, mannan, galactan, and starch; moreover, some members of a few CBM families can recognize mammalian glycans (6). CBMs target enzymes to their specific substrates, enhancing carbohydrate degradation as a result of the increased local concentration of each enzyme around its substrate (4, 7). The frequent occurrence of CBMs with specificity for cellulose in enzymes, such as xylanases, without activity on this polysaccharide, can be explained by the close proximity of these two carbohydrates, cellulose and xylan, in cell wall structure in biomass (7). In a similar way, CBMs of scaffolding proteins play an important role in the deconstruction of the plant cell wall by keeping the clustered catalytic components of cellulosomes in close proximity to the cell wall, yielding an enhanced and synergistic degradation of polysaccharides (8, 9). A previous proposal that CBMs may act by a disruptive mechanism that separates glycan chains from the cell wall (10) was supported by the finding of some CBM33s, not attached to enzymes, which disrupted chitin structure and promoted hydrolysis by chitinases (11). However, it has been recently shown that CBM33s constitute a new type of carbohydrate depolymerases that act by an oxidative mechanism and are reclassified in CAZy as auxiliary activity family 10 (12). Furthermore, the analysis of some CBM35s showed that they display specificity for Δ 4,5-anhydrogalacturonic acid, a signature molecule of plant cell wall degradation by pectate lyases, suggesting that CBMs of this family could direct the enzymes toward regions of the plant cell wall that are being actively degraded or remodeled (13). In addition to targeting the catalytic modules to their substrates, CBMs probably have a more complex role in polysaccharides depolymerization, with additional functions to enable the deconstruction of the cell wall and the catalytic degradation of carbohydrates.

CBMs of several families show specificity for xylan (2, 14–16). Among them, family 22 CBMs were previously considered as thermostabilizing domains of enzymes from thermophiles (17). The finding that they were not restricted to thermo-

philic enzymes prompted the analysis of their binding properties, showing that they promote binding to xylan and xylo-oligosaccharides (18, 19). They were regarded as xylan binding domains that were reclassified as the new CBM22 (20, 21). Xylanase Xyn10C belongs to the secretome of *Paenibacillus barcinonensis*, a powerful xylanolytic microorganism (22) that produces a complex set of enzymes, including xylanases of families GH10, GH11, and GH30 that have been cloned and characterized (16, 23, 24). Xyn10C was previously characterized showing a distinctive modular structure containing an N-terminal tandem of two CBM22s and a duplicated CMB9 at its C terminus (25). In our study, we have solved the three-dimensional structure of the CBM22-1–CBM22-2 tandem, and its binding abilities have been investigated by soaking experiments, affinity gel electrophoresis, and ITC. In addition, the structure and binding features of the individual CBM22-1, CBM22-2, and the catalytic GH10 modules are reported. The results found contribute to deciphering the function of CBM22s and their contribution to xylan degradation. Further studies will be required to fully ascertain the role of CBM22 duplicity in catalytic depolymerization of xylan by their partner enzymes and in deconstruction of biomass.

Experimental Procedures

Cloning, Expression, and Purification—The respective forward and reverse primers used for all constructs are given in Table 1. The cDNA of Xyn10C was amplified from *P. barcinonensis* BP-23 genomic DNA (UniProt accession number O69230). The resulting plasmid pET28Xyn10C produced the full-length enzyme linked to an N-terminal His₆ tag (Xyn10C). To maintain the His tag motif in the N terminus of the expressed protein, the first 28 residues of the sequence were not included in the DNA amplification because they correspond to the signal peptide that is cleaved in *Escherichia coli* hosts during protein translation. The plasmid vector pET28 was used to clone the GH10 catalytic domain fused to an N-terminal His₆ tag (Xyn10C-CD). Both constructs (Xyn10C and Xyn10C-CD) were expressed and purified using a similar protocol. BLR (DE3) *E. coli* strain was transformed with the corresponding plasmid and grown in LB medium at 37 °C for 4 h until the A_{600} reached 0.8. Protein expression was induced by adding 0.3 mM isopropyl β -D-thiogalactopyranoside followed by incubation for 15 h at 16 °C. Cells were disrupted by using a French press. The recombinant His₆ tag proteins were purified from cell extracts by immobilized metal affinity chromatography using HisTrap HP columns of 5 ml (GE Healthcare) and eluted in 20

Structure and Function of Xyn10C

mM sodium phosphate buffer, pH 7.4, 500 ml of NaCl with a 0–500 mM imidazole gradient. For crystallization, it was necessary to cleave the His₆ tag of Xyn10C-CD with thrombin (Sigma).

The regions encoding for the N-terminal domain, Xyn10C-XBD, and the CBM22-1 and CBM22-2 modules were also amplified and cloned into the pGEX-4T-2 vector (GE Healthcare). The GST-fused protein expression was performed under similar conditions to those described above, except the BL21 (DE3) *E. coli* strain was employed. All constructs were purified using a similar protocol. Cell pellets were resuspended in buffer A (20 mM Tris/HCl, pH 8, 100 mM NaCl) and disrupted through sonication. The clarified cell lysates were mixed with 10 ml of glutathione-Sepharose 4B resin (GE Healthcare) equilibrated in buffer A. After washing with buffer A, the target protein was cleaved from resin-bound GST by incubating with thrombin overnight. Eluted protein was diluted to 50 mM NaCl and loaded to a 5-ml anionic exchange column (HiTrap Q HP; GE Healthcare) equilibrated in buffer B (50 mM NaCl, 20 mM Tris/HCl, pH 8). After washing with buffer B, the protein was eluted using a gradient between 50 and 500 mM NaCl. In the case of Xyn10C-XBD, the protein was further purified by size exclusion chromatography using a 16/60 Superdex 200 column (GE Healthcare) equilibrated in buffer A. The protein purity was assessed by SDS-PAGE in all cases. Pure proteins were concentrated and frozen at –80 °C until use.

Binding to Insoluble Polysaccharides—Binding activity to insoluble polysaccharides was assessed as described by Hogg *et al.* (26) with some modifications. Briefly, 250 µg of purified Xyn10C, Xyn10C-XBD, or Xyn10C-CD was mixed with 25 mg of Avicel or insoluble oat spelt xylan in a final volume of 500 µl of 50 mM Tris, pH 7, in 1.5-ml microcentrifuge tubes. The samples were incubated for 1 h at 4 °C with gentle orbital mixing. Samples were then centrifuged at 18,000 × g for 20 min, and supernatants, containing unbound protein, were carefully removed. Pellets were washed three times with 400 µl of the same buffer, before being resuspended in 400 µl of 10% SDS and boiled for 10 min to release bound protein. Samples were then analyzed by SDS-PAGE on 8 or 12% (w/v) polyacrylamide gels.

Affinity Gel Electrophoresis and Zymogram Analysis—Affinity gel electrophoresis was performed by following the method of Correia *et al.* (27). Continuous native polyacrylamide gels containing 10% acrylamide in 25 mM Tris, 250 mM glycine buffer, pH 8.3, were used. Soluble xylan (4 mg/ml) was included in gels before polymerization. Gels, with and without xylan, were polymerized at the same time and run in the same gel tank. About 6 µg of target protein was loaded in each well at room temperature, and gels were run at 10 mA/gel for 2 h. Bovine serum albumin (BSA) was used as a negative noninteracting control.

To analyze the electrophoretic homogeneity of purified proteins, SDS-PAGE was performed in 15% (w/v) (for CBM22-1 and CBM22-2) and 12% (w/v) (for Xyn10C-XBD) polyacrylamide gels essentially as described by Laemmli (28). For detection of xylanase activity, zymogram analysis was performed in SDS-polyacrylamide gels containing 0.2% (w/v) birchwood xylan. Samples were heated for 10 min at 45 °C in sample buffer before being applied to gels. After electrophoresis, gels were

soaked in 2.5% (w/v) Triton X-100 for 30 min, washed in 50 mM acetate buffer, pH 5.0, for 30 min, and incubated at 45 °C for 2 h in the same buffer. Gels were then stained with 0.1% (w/v) Congo red for 15 min and washed with 1 M NaCl until xylanase bands became visible, as described previously (29). Gels were then immersed in 10% (v/v) acetic acid and photographed.

Isothermal Titration Calorimetry—ITC measurements were performed at 25 °C using a MicroCal VP-ITC (GE Healthcare). The samples were dialyzed against 50 mM sodium phosphate, pH 7, and ligand solutions were prepared using the same buffer. All samples were thoroughly degassed before use. Titrations were carried out by injecting consecutive aliquots of 10–20 mM oligosaccharide or 4 mg/ml xylan oat spelt polysaccharide into the sample cell loaded with protein at 50–65 µM. For mixed oligosaccharide and xylan oat spelt, two consecutive titrations were carried out with the same protein solution, and the CONCAT32 program (GE Healthcare) was used to concatenate the raw data of both series. Heat developed on ligand dilution was determined separately and subtracted when required. The binding constants (*K*) and the enthalpies (ΔH) of binding were calculated by analyzing the binding isotherms with the MicroCal ITC Origin software, assuming one carbohydrate-binding site per domain, based on the structural data and xylan binding capacity of the isolated domains. The entropic contribution to ligand binding was calculated using Gibbs equation ($\Delta G = -RT \ln K = \Delta H - T\Delta S$). The molar concentration of binding sites in the polysaccharide was set to the value that after iteration gave a number of sites per CBM22 domain equal to 1. The oligosaccharides were purchased from Megazyme. The mixed oligosaccharides used were 1,3:1,4- β -glucotetraose A (Glc-3Glc-4Glc-4Glc), B (Glc-4Glc-4Glc-3Glc), and C (Glc-4Glc-3Glc-4Glc).

Analytical Ultracentrifugation—Sedimentation velocity experiments were run at 45,000 rpm using cells with double sector Epon-charcoal centerpieces. Differential sedimentation coefficients were calculated by least squares boundary modeling of the experimental data with the program SEDFIT. Sedimentation equilibrium experiments were carried out at different rotor speeds, as described previously (30), and the weight-average molecular weights were calculated using the heteroanalysis program (Biotechnology-Bioservices Center, University of Connecticut). Measurements were performed at 20 °C in an Optima XL-A analytical ultracentrifuge (Beckman Coulter) using Xyn10C-XBD (5.3–24.3 µM) or CBM22-1 (17.3 µM) in ITC buffer (with and without 3 mM xylotetraose). Partial specific volumes were calculated from the amino acid sequences with the SEDNTERP program.

Crystallization and Data Collection—Crystals of Xyn10C-XBD were grown from 47 mg/ml protein, 25% (w/v) PEG 3350, 0.2 M NaCl, 2% (v/v) 1,6-hexanediol, 0.1 M BisTris, pH 5.5, using the streak-seeding technique. Very thin plates grew by vapor diffusion at room temperature, as described before (31). For data collection, crystals were transferred to cryoprotectant solutions consisting of mother liquor plus 20% (v/v) glycerol before being cooled in liquid nitrogen.

Screening of crystallization conditions for the isolated CBM 22s and Xyn10C-CD domains was performed with the PACT and the JCSG Suites from Qiagen and INDEX and SALTRX

TABLE 2
Crystallographic data

Values in parentheses are for the high resolution shell.

| Crystal data | Xyn10C-XBD | CBM22-1 | CBM22-2 | CBM22-2/Xyl4 | CBM22-2/Xyl3 | CBM22-2/Glc-4 Glc-4Glc-3Glc | Xyn10C-CD |
|--|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|--------------------------------|---------------------------|
| Space group | <i>P</i> 2 ₁ | <i>H</i> 3 | <i>P</i> 3 ₂ | <i>P</i> 3 ₂ | <i>P</i> 3 ₂ | <i>P</i> 3 ₂ | <i>P</i> 3 ₂ |
| Unit cell | | | | | | | |
| <i>a</i> (Å) | 84.17 | 85.35 | 92.54 | 92.77 | 92.46 | 92.48 | 126.69 |
| <i>b</i> (Å) | 110.36 | 85.35 | 92.54 | 92.77 | 92.46 | 92.48 | 126.69 |
| <i>c</i> (Å) | 118.52 | 108.46 | 48.40 | 48.57 | 48.43 | 48.38 | 57.70 |
| γ (°) | 90.63 | | | | | | |
| Data collection | | | | | | | |
| Beamline | ID29(ESRF) | ID23-2(ESRF) | ID23-2(ESRF) | ID23-2(ESRF) | BL13-XALOC (ALBA) | BL13-XALOC (ALBA) | ID23-1 (ESRF) |
| Wavelength (Å) | 0.979235 | 0.8729 | 0.8729 | 0.8729 | 0.9795 | 0.9795 | 0.9801 |
| Resolution (Å) | 84.16–2.43 (2.48–2.43) | 30.54–1.70 (1.79–1.70) | 33.45–1.75 (1.84–1.75) | 33.55–1.67 (1.76–1.67) | 48.42–1.95 (2.05–1.95) | 48.38–1.80 (1.90–1.80) | 36.57–1.76 (1.86–1.76) |
| Data processing | | | | | | | |
| Total reflections | 510618 (29165) | 184811 (26283) | 272776 (39214) | 307899 (41596) | 315365 (44621) | 393762 (54642) | 567570 (81227) |
| Unique reflections | 81116 (4468) | 32396 (4730) | 46784 (6810) | 54496 (7920) | 33856 (4948) | 42863 (6228) | 102672 (14987) |
| Multiplicity | 6.3 (6.5) | 5.7 (5.6) | 5.8 (5.8) | 5.6 (5.3) | 9.3 (9.0) | 9.2 (8.8) | 5.5 (5.4) |
| Completeness (%) | 99.4 (99.3) | 100 (100) | 100 (100) | 100 (100) | 99.9 (99.7) | 100 (100) | 100.0 (100.0) |
| <i>I</i> / σ (<i>I</i>) | 4.8 (1.3) | 2.4 (1.7) | 5.8 (1.9) | 6.1 (1.7) | 6.2 (2.5) | 5.8 (1.7) | 5.1 (1.4) |
| Mean <i>I</i> / σ (<i>I</i>) | 8.6 (2.2) | 9.7 (4.2) | 13.4 (4.4) | 13.1 (3.4) | 17.5 (7.6) | 14.8 (4.6) | 9.1 (3.2) |
| <i>R</i> _{merge} ^a (%) | 13.0 (65.2) | 13.5 (41.6) | 9.4 (40.3) | 9.1 (45.1) | 9.0 (28.7) | 9.2 (43.2) | 10.6 (49.4) |
| <i>R</i> _{pim} ^b (%) | 8.5 (34.9) | 6.2 (19.3) | 4.3 (18.4) | 4.2 (21.8) | 3.1 (10.1) | 3.2 (15.3) | 5.0 (23.5) |
| Molecules per a.u. | 6 | 2 | 3 | 3 | 3 | 3 | 3 |
| Matthews coefficient (Å ³ Da ^{−1}) | 2.54 | 2.24 | 2.23 | 2.25 | 2.23 | 2.23 | 2.26 |
| Solvent content (%) | 51.5 | 45.2 | 44.9 | 45.0 | 44.8 | 44.8 | 45.5 |
| Overall <i>B</i> -factor Wilson plot (Å ²) | 37.24 | 18.79 | 12.64 | 11.79 | 16.39 | 18.16 | 15.55 |
| Refinement | | | | | | | |
| <i>R</i> _{work} / <i>R</i> _{free} ^c (%) | 22.52/26.05 | 21.09/24.71 | 15.14/18.20 | 15.78/19.37 | 17.92/21.60 | 20.41/22.57 | 21.79/25.15 |
| No. of atoms/average <i>B</i> (Å ²) | | | | | | | |
| Protein | 12529/49.18 | 2398/25.29 | 3789/17.74 | 3789/16.40 | 3789/21.39 | 3789/24.20 | 8355/16.55 |
| Carbohydrate | 0 | 0 | 0 | 101/26.22 | 66/33.23 | 23/41.73 | 0 |
| Other | 22/41.13 | 2/19.9 | 3/16.69 | 3/18.41 | 3/20.21 | 3/23.53 | 21/23.63 |
| Water molecules | 29/30.30 | 59/24.16 | 231/21.69 | 94/15.76 | 84/19.39 | 113/23.45 | 523/21.60 |
| Ramachandran | | | | | | | |
| Favored (%) | 98.00 | 98.05 | 97.69 | 97.48 | 97.48 | 97.48 | 95.43 |
| Outliers (%) | 0.00 | 0.32 | 0.00 | 0.00 | 0.00 | 0.00 | 0.38 |
| r.m.s.d. | | | | | | | |
| Bonds (Å) | 0.0108 | 0.0106 | 0.0123 | 0.0190 | 0.0153 | 0.0146 | 0.007 |
| Angles (°) | 1.374 | 1.405 | 1.451 | 1.622 | 1.425 | 1.593 | 1.147 |
| Protein Data Bank codes | 4XUP | 4XUO | 4XUN | 4XUR | 4XUQ | 4XUT | 4W8L |

^a $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th measurement of reflection *hkl* and $\langle I(hkl) \rangle$ is the weighted mean of all measurements.^b $R_{\text{pim}} = \sum_{hkl} (1/(N-1)) |I(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} I(hkl)$, where *N* is the redundancy for the *hkl* reflection.^c $R_{\text{work}}/R_{\text{free}} = \sum_{hkl} F_o - F_c / \sum_{hkl} F_o$, where F_c is the calculated and F_o is the observed structure factor amplitude of reflection *hkl* for the working/free (5%) set, respectively.

from Hampton Research. Crystals from CBM22-1 appeared in solutions containing either PEG 1500 or PEG 3350 with pH 5.5–7.0. Further optimization was performed at room temperature and the vapor diffusion method by mixing 29 mg/ml protein sample (20 mM Tris, pH 8.0, and 125 mM NaCl) and an equal volume of different reservoir solutions in a sitting-drop procedure. Large crystals grew from 20% (w/v) PEG 3350 and 0.2 M NaNO₃, with a 1.5:1 ratio of protein/precipitant, reaching maximum size in 10 days. Treatment with different cryoprotectants as glycerol, PEG 400, or xylose seemed to greatly affect the crystal integrity, with only a very fast soaking into 15% (v/v) ethylene glycol being tolerated and thus allowing collection of high quality data (Table 2). Soaking with ligand solutions damaged crystals, and co-crystallization experiments were unsuccessful, therefore precluding analysis of complexes.

Rod-shaped crystals from the CBM22-2 construct were obtained by mixing 0.5 μ l of 19 mg/ml protein solution (20 mM Tris, pH 8.0, and 150 mM NaCl) with 1 μ l of a solution containing 1.85 M sodium malonate, pH 6.0, and equilibrating by vapor

diffusion at room temperature. Complexes were obtained by the soaking technique using xylotriose, xylotetraose, or 1,3:1,4 β -glucotetraose B. The crystals were soaked for 5–30 min in solutions made of mother liquor plus 10 mM of the corresponding ligand. For data collection, all crystals were transferred to cryoprotectant solutions containing 3 M sodium malonate before being cooled in liquid nitrogen.

Rod-shaped crystals from Xyn10C-CD were grown by mixing equal amounts of 10 mg/ml protein solution (in 20 mM Tris, pH 7.5, 0.5 M NaCl) with 23% (w/v) PEG 3350, 0.2 M KSCN, 0.1 M NaAc, pH 5.0, and equilibrating at room temperature. For data collection, 20% glycerol was added to mother liquor before cooling.

X-ray data from all crystals were collected using different Synchrotron sources at Alba (Spain) and ESRF (France) on the beamlines given in Table 2. Diffraction images were processed with iMOSFLM (32) or XDS (33) and merged using the CCP4 package (34).

Structure Solution and Refinement—The structure of the Xyn10C-XBD tandem was solved by molecular replacement

Structure and Function of Xyn10C

using the PHASER program (35). The search models were the CBM22-1 and CBM22-2 domains from the *Ruminococcus thermocellum* xylanase 10B (RthXyn10B; PDB codes 2W5F and 1DYO, (20, 36)), each sharing 21/25 and 23/23% sequence identity, respectively, with respect to the CBM22-1 and CBM22-2 domains. Nine units of the search models were placed in the asymmetric unit by PHASER. Then an additional unit was manually included on the basis of the observed association pattern, which increased the likelihood score and showed appropriate packing. The solution means six molecules of the Xyn10C-XBD tandem within the asymmetric unit (a.u.), with two of them being incomplete. This initial model was subjected to restrained refinement with REFMAC (37), within the CCP4 suite, with flat bulk-solvent correction and local noncrystallographic symmetry. Then, the calculated electron density maps allowed unambiguous assignment of the CBM22-1 and CBM22-2 sequences. Further refinement with REFMAC, combined with manual building using the program COOT (38), led to a model showing an almost continuous density for the whole polypeptide chain in four molecules (residues 1–334), a small fragment of the linker region showing poor or chopped density as follows: A (residues 164–166), B (residues 162–171), D (residues 161–170), and F (residues 165–170), whereas only the chain comprising residues 3–162 could be modeled in molecules D and F. Many attempts to complete this model by manual and automatic rebuilding were unsuccessful. At the later stages, water molecules were included, which, combined with more rounds of restrained refinement, led to a final *R*-factor of 22.7 (*R*_{free} = 26.0) for all data set up to 2.43 Å resolution. Refinement parameters are reported in Table 2.

The structure of CBM22-1 and CBM22-2 crystals was solved by molecular replacement using MOLREP (39) and the coordinates of the 1–159 and 171–334 portion of Xyn10C-XBD, respectively, as the search models. Crystallographic refinement was performed using REFMAC (37) combined with model building with COOT (38) and the addition of water molecules, which led to the final *R*-factors given in Table 2. The structures of the CBM22-2 complexes were solved by difference Fourier synthesis using the refined coordinates of the CBM22-2 domain. The ligands were manually built in the electron density maps and were refined similarly, to reach the *R*-factors listed in Table 2.

The structure of XynC-CD was solved by molecular replacement using MOLREP (39) and the intracellular xylanase from *Geobacillus stearothermophilus* (PDB code 2Q8X) as search model. The model was refined with REFMAC (37) and completed with COOT (38). The crystals belong to the P₃₂ space group with three independent molecules within the a.u. that are essentially identical, showing root mean square deviation (r.m.s.d.) in the range 0.17–0.24 Å when superimposing their 352 α -carbon atoms C α s.

Stereochemistry of the models was checked with PROCHECK (40) and MOLPROBITY (41). The figures were generated with PyMOL (42). r.m.s.d. analyses were made using the program SUPERPOSE within the CCP4 package (34). The interfaces were analyzed by PISA (43).

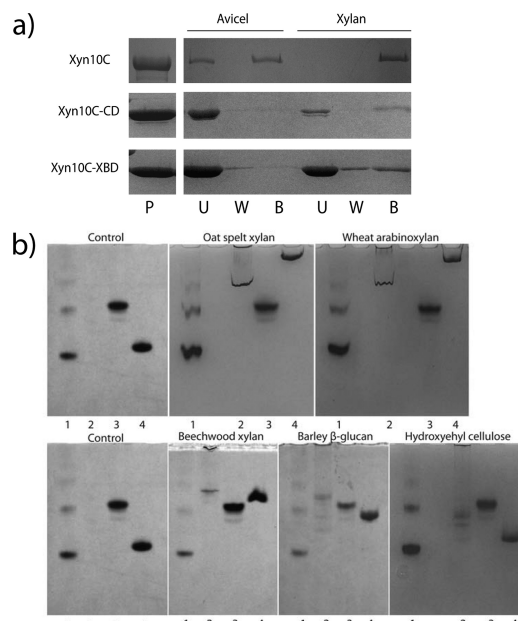


FIGURE 1. Binding specificity of Xyn10C. a, SDS-PAGE analysis of binding to insoluble polysaccharides. Proteins (P) were mixed with Avicel or with the insoluble fraction of oat spelt xylan for 1 h; unbound (U) and bound (B) fractions were separated by centrifugation; pellets were washed three times (W). b, nondenaturing polyacrylamide gels containing no ligand (control) or soluble polysaccharides. Lane 1, BSA; lane 2, Xyn10C; lane 3, Xyn10C-CD; lane 4, Xyn10C-XBD.

Results

N-terminal Xyn10C Region Binds Xylan—Specificity of Xyn10C binding to insoluble polysaccharides was investigated. Avicel and insoluble oat spelt xylan were incubated with the complete Xyn10C and its isolated Xyn10C-XBD and Xyn10C-CD domains, and their bound and unbound fractions were separated by centrifugation and analyzed by SDS-PAGE (Fig. 1a). Binding to Avicel is observed only in the complete enzyme, whereas the isolated Xyn10C-XBD and Xyn10C-CD domains only bind insoluble xylan, apparently less efficiently than the full-length protein. Consequently, only the C-terminal CBM9–1–CBM9–2 region has cellulose-binding ability.

Binding of the three samples to soluble fractions of xylan (wheat arabinoxylan, oat spelt xylan, and beechwood xylan), barley β -glucan, and hydroxyethylcellulose was also evaluated using affinity gel electrophoresis (Fig. 1b). Migration of full-length Xyn10C and Xyn10C-XBD in gels containing soluble xyans and barley β -glucan was markedly retarded with respect to migration in gels without the substrates, whereas no binding to hydroxyethylcellulose was observed by any of them. Arabinoxylans from oat spelt and wheat produced the largest effect, whereas beechwood xylan and barley β -glucan show a comparable and smaller effect. Therefore, Xyn10C and its Xyn10C-XBD region present higher affinity for arabinoxylans than for glucuronic-substituted xylan or mixed β -glucans. In contrast,

Structure and Function of Xyn10C

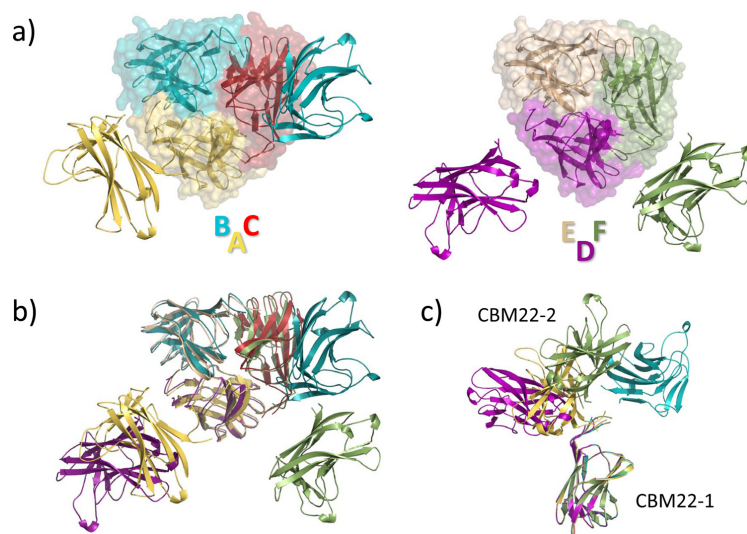


FIGURE 2. **Crystal structure of the Xyn10C-XBD tandem.** *a*, asymmetric unit contains six independent molecules forming two trimers, ABC (left) and DEF (right); CBM22-1 domains are highlighted with a transparent surface. Molecules C and E could not be full-traced in the electron density maps. *b*, superposition of the two trimers showing that both are arranged by a common association pattern of their CBM22-1 moieties, whereas the second CBM22-2 domain is asymmetrically located. *c*, different conformations of the tandem observed in the crystals, with the linker being partially disordered.

migration of Xyn10C-CD domain was not retarded by any substrate.

Xyn10C-XBD Crystal Structure Shows a Multimeric Arrangement—The crystallization of a construct containing the xylan-binding N-terminal portion (residues 1–334) of xylanase 10C from *P. barcinonensis* (Xyn10C-XBD) has been previously reported (31). We present here its three-dimensional structure determined by molecular replacement at 2.4 Å resolution. Details are given under “Experimental Procedures” and in Table 2. The structure of Xyn10C-XBD reveals the presence of two separate domains, corresponding to CBM22-1 (residues 1–159) and CBM22-2 (residues 171–334) repeats, both adopting a β -sandwich fold, as explained below.

The final model contains six Xyn10C-XBD molecules in the a.u., although two of them could not be fully traced. The six molecules are arranged in two trimers, ABC and DEF, and the CBM22-2 portion of subunits C and E was not visible in the electron density maps (Fig. 2*a*). However, structural superimposition of the two observed trimers allows the depiction of a putative “full” trimer that would be represented by molecules A/DBF (Fig. 2*b*). Crystal packing shows cavities where the CBM22-2 moiety of molecules C and E could be allocated, but the electron density maps only showed residual noise despite the many attempts to perform manual and automatic building. Taking into account that mass spectrometry excludes proteolysis of the polypeptide chain, we assume that the missing parts are disordered in the crystals.

As shown in Fig. 2*a*, both trimers lack noncrystallographic 3-fold symmetry. *i.e.* each protomer adopts a different molecular conformation (Fig. 2*c*) within the trimers. However, it should be noted that each trimer is built up from a regular association of three CBM22-1 domains (Fig. 2*a*) that bury 700

Å² (*p* value 0.59). In contrast, the observed asymmetrical orientation of the corresponding CBM22-2 moieties is probably determined by the highly intricate crystal packing interactions. The association pattern observed in Xyn10C-XBD is commented on below.

Two Differing CBM22 Modules Are Attached with a Great Flexibility—The main consequence of these results is the high flexibility presented by the N-terminal domain of Xyn10C. Thus, the structural comparison of the four full-traced molecules, illustrated in Fig. 2*d*, reveals a broad molecular conformational landscape in which CBM22-2 may adopt different dispositions with respect to CBM22-1. This flexibility is allowed by the long proline-glycine-rich ¹⁶⁰PENPGEPGEAG¹⁷⁰ segment linking both modules, which is partially disordered in the crystal (a segment varying from 4 to 10 residues, depending on the different molecules within the a.u., is unobserved) and must provide high mobility to the polypeptide chain.

Each CBM22 domain is folded into a classical lectin-like β -jelly roll composed, mainly, of two antiparallel β -sheets, each with six (β_9 , β_6 , β_{11} , β_1 , β_4 , and β_3) and five β -strands (β_8 , β_7 , β_{10} , β_5 , and β_2) (Fig. 3, *a* and *b*). One of these β -sheets shapes an extended cleft encompassing the three aromatic residues from the substrate-binding motif, Trp-55/Tyr-217, Tyr-104/Tyr-268, and Tyr-136/Tyr-303 (in CBM22-1/CBM22-2 numbering), the two first contouring a sandwiched hydrophobic platform (Fig. 3, *b* and *c*). The cleft also presents two conserved residues, an arginine (Arg-27/Arg-189) and a glutamate (Glu-138/Glu-305) that have been proved to be absolutely required for xylan recognition in CBM22-2 from *RthXyn10b* (21). Structural comparison of CBM22-1 and CBM22-2 shown in Fig. 3*b* reveals an r.m.s.d. of 1.9 Å, based on superposition of 146 C α atoms. Despite their rather modest sequence similarity (39%)

Structure and Function of Xyn10C

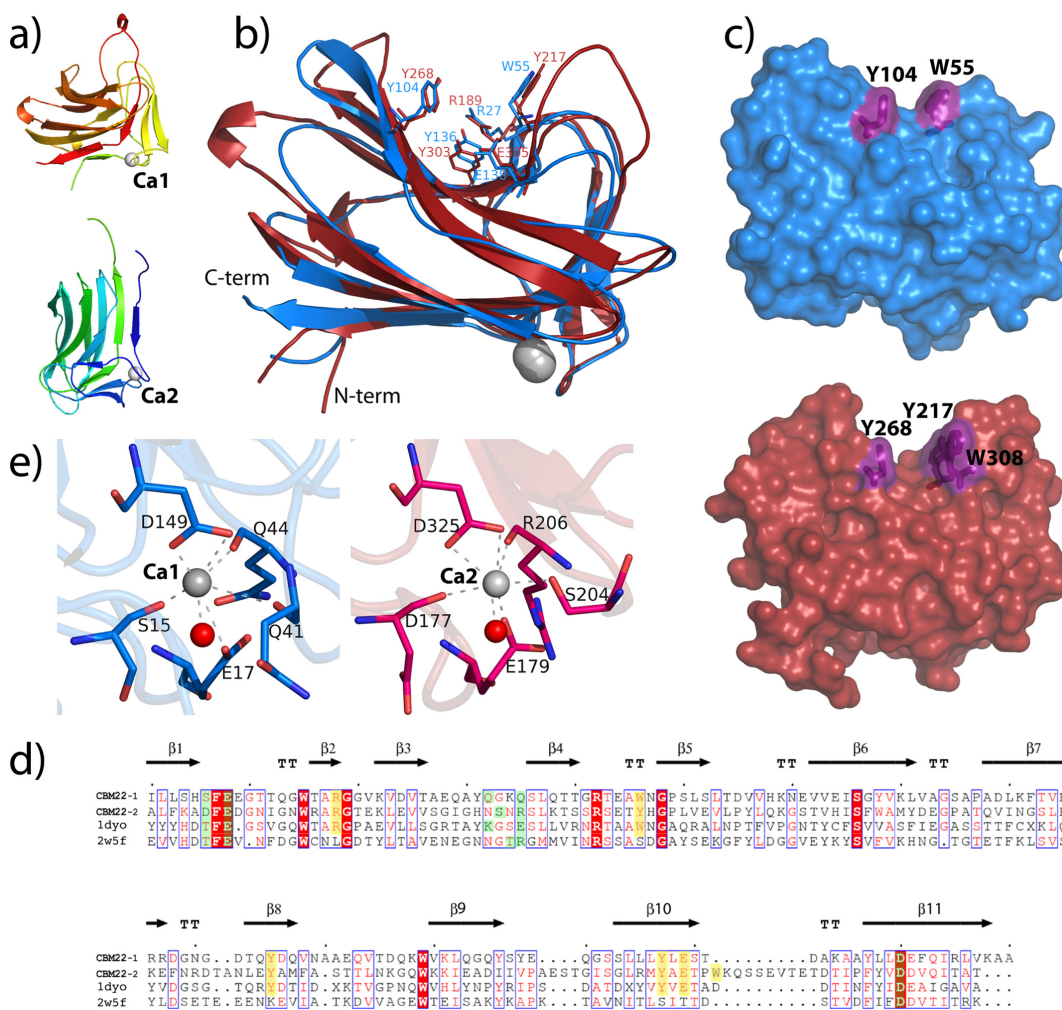


FIGURE 3. Topology of the CBM22-1 and CBM22-2 domains. *a*, secondary structure of the Xyn10C-XBD tandem in rainbow sequence code. The two calcium ions are shown as white spheres. *b*, superimposition of CBM22-1 (blue) onto CBM22-2 (red) showing the five residues previously reported to be required for xylan binding. Topological differences between both domains are concentrated at some loops surrounding the binding cleft. *c*, molecular surface of CBM22-1 (top) and CBM22-2 (bottom), highlighting the hydrophobic platform defined by the aromatic residues located at the binding cleft. *d*, structure-based sequence alignment of CBM22-1 and CBM22-2 with the CBM22-1 (PDB code 2W5F) and CBM22-2 (PDB code 1DYO) modules of *RthXyn10B* (20, 36). Residues involved in ligand binding are highlighted in yellow, and residues coordinating the calcium ion are marked in green. *e*, zoom of the calcium ion-binding site of CBM22-1 (blue) and CBM22-2 (red) showing the coordination pattern at each domain.

and identity (23%), the topology of the two domains is essentially conserved. The main differences are located in some loops on the concave face of the β -sandwich, which are longer in the CBM22-2 domain; these are Arg-261–Leu-266, at the end of strand $\beta 7$, and the loops Pro-291–Ser-298 and Trp-308–Glu-316, before and after strand $\beta 10$ (Fig. 3*d*). It is known that differences in these loops may alter radically the substrate affinity on analogous CBMs, and in the case of Xyn10C-XBD, they generate a significantly deeper groove in CBM22-2. Particularly remarkable is the long loop Trp-308–Glu-316 inserted at the

end of strand $\beta 10$ that contains Trp-308, an additional aromatic residue next to Tyr-217 that is solvent-exposed and extends the hydrophobic platform, shaping new subsites at the binding cleft (Fig. 3*c*) as described below. Finally, and in common with most β -sandwich CBMs, there is a conserved calcium ion having a structural role, located on the surface at the opposite face of the groove. This calcium ion is coordinated by the carboxylates of Glu-17/Glu-179 and Asp-149/Asp-325, both conserved within the CBM22 family (Fig. 3*e*). The coordination sphere is completed with the main chain carbonyls from Ser-15,

Structure and Function of Xyn10C

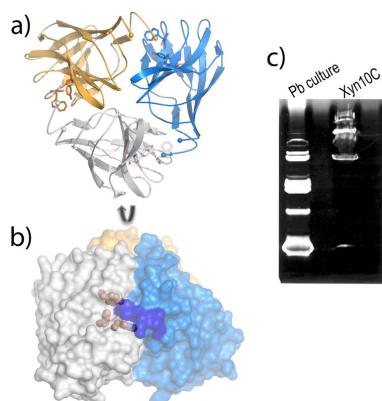


FIGURE 4. Crystal structure of the isolated CBM22-1 domain. *a*, packing of each independent molecule within the a.u. shows a trimeric association, where the subunits are colored in blue, gold, and white. *b*, perpendicular view of this trimer represented as molecular surface, showing the aromatic residues from the binding motif as gold spheres, and the region occupied by the N-terminal segment of the adjacent subunit highlighted in blue. *c*, zymogram analysis of Xyn10C; 1st lane, xylanases identified from a *P. barcinonensis* culture; 2nd lane, recombinant sample of Xyn10C showing the presence of different active forms.

Gln-41, and Gln-44, in domain CBM22-1, and by the carbonyls from Asp-177 and Arg-206, and the side chain of Ser-204, in the case of domain CBM22-2. Both calcium ions are also coordinated to a water molecule configuring an octahedral arrangement.

Because of the difficulties in handling the very thin and fragile plates, and the limited resolution shown by many of them, each domain was expressed independently with the aim to undertake soaking experiments. However, only the CBM22-2 crystals provided suitable complexes to investigate its binding specificity, as explained below.

Isolated CBM22-1 Domain Keeps in the Crystal the Same Association Trend of the Tandem—A construct encoding the N-terminal CBM22-1 domain was produced, and its structure has been determined at 1.7 Å resolution (Table 2). The crystals belong to the H3 space group, with two molecules in the a.u., A and B, which are essentially equivalent, with an r.m.s.d. of 0.172 Å after superposition of 135 Cα atoms. The electron density maps showed the full polypeptide chain, including Ala-4–Ala-159. Interestingly, analysis of the structure of the CBM22-1 crystals shows an identical association pattern to that described in the Xyn10C-XBD crystals, *i.e.* crystal packing of each of the two independent molecules A/B generates a similar trimer (Fig. 4*a*) that is equivalent to that presented by the tandem construct (Fig. 2*a*). Thus, the CBM22-1 domains from the ABC tandem trimer superimposes onto the AAA trimers with an r.m.s.d. of 0.673 Å for 410 Cα atoms. Interestingly, the number of polar interactions between subunits is moderate (eight H-bonds and three salt-bridges), but the number of van der Waals contacts (56 contacts shorter than 4 Å), and the evident shape complementarity among subunits is remarkable. Furthermore, an inspection of the CBM22-1 trimer interface, shown in Fig. 4*b*, reveals that the aromatic residues from the substrate-binding motif are located close to the interface, and the N-terminal

segment from the adjacent subunits is filling the binding cleft. Consequently, diffusion of the ligands into the binding site probably disrupts the interface, which might be deleterious for crystal packing. This might be the reason why soaking experiments with ligands dislocates most crystals or led to highly twinned data, precluding the structural characterization of the complex.

The physiological significance of a putative trimeric association of the CBM22-1 domain is unclear. The association buries only 21% of the total molecular surface (4209/20,129 Å²), and apparently, the trimer formation might constrain the binding ability of CBM22-1 due to the proximity of the binding cleft to the interface. In fact, only terminal xylooligosaccharides could be allocated within each binding cleft in the trimer. However, the absolute conservation of this assembly through Xyn10C-XBD and its isolated CBM22-1 domain crystals is remarkable. The outstanding feature is that the association trait is conserved in the CBM22-1 crystals in the absence of the CBM22-2 moiety, which was doing many crystal contacts in the tandem. This observation might suggest that the association trend observed in both crystals might be a trait of the CBM22-1 domain. Although ultracentrifugal analysis has shown that both the Xyn10C-XBD tandem and the CBM22-1 domain are primarily monomeric in solution (data not shown), active oligomeric forms have been detected by zymogram analysis of full-length Xyn10C (Fig. 4*c*). Whether these observations assign a role to CBM22-1 in oligomerization of Xyn10C requires further investigation.

CBM22-2 Complexes Illustrate Its Ligand-binding Mode—A construct encoding the truncated CBM22-2 domain, which encompasses residues 172–332, was produced, and its structure has been determined at 1.75 Å resolution. Crystals belong to the P3₂ space group with three molecules within the a.u. that do not associate and are essentially equivalent (r.m.s.d. 0.281/0.318 Å for 143/145 Cα atoms when superimposing molecules B or C, respectively, to A). The only significant differences are observed at loop 308–313, within the insertion located at the end of β10, which is involved in substrate binding, as explained below.

To investigate the molecular basis for ligand specificity, soaking experiments of the crystals were performed with xylotriose, xylotetraose, and 1,3:1,4 β-glucotetraose B. As shown in Fig. 3*c*, CBM22-2 contains a deep cleft, 20 Å length, 9 Å deep, running across the concave surface of the β-sheet that conforms the substrate-binding site. In the middle of this cleft, three protruding aromatic residues define a sandwich hydrophobic platform to allocate the substrate (Fig. 5). The side chains of Tyr-217 and Trp-308, at one face of the platform, are making an angle of 140°, thereby shaping a somehow twisted configuration for ligand binding. The electron density of the xylotetraose-soaked crystals show four ordered xylose units within this binding cleft (Fig. 5*a*), which have been ascribed to subsites S1 to S4 going from the reducing to the nonreducing end, and covers the whole surface along the crevice. The direction of the sugar chain was assigned by evaluating the temperature factors of the two alternative O5/C5 positions and the hydrogen-bonding pattern of O5. The xylooligosaccharide binds in the predicted 3-fold helical conformation previously determined by x-ray

Structure and Function of Xyn10C

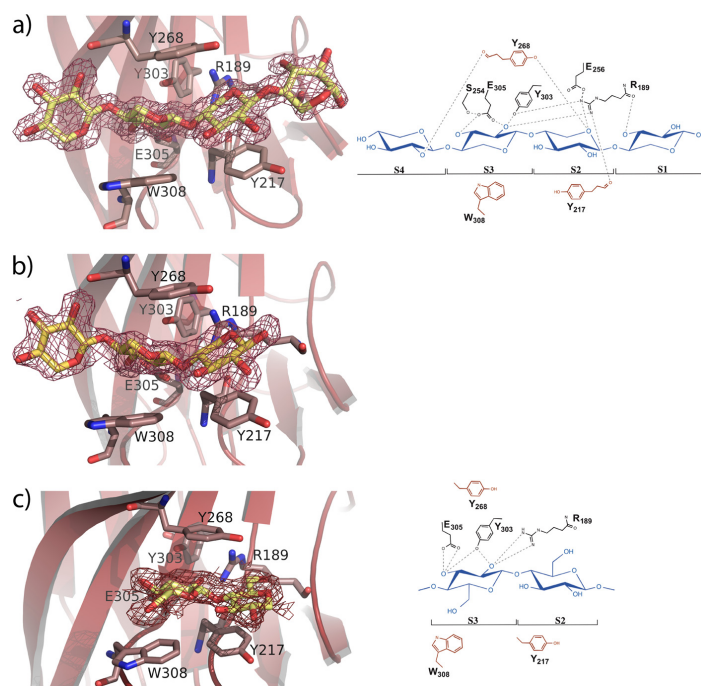


FIGURE 5. **Structure of the CBM22-2 complexes.** *a*, xylotetraose; *b*, xylotriose; *c*, the mixed glucan Glc-4Glc-4Glc-3Glc, bound at the CBM22-2 binding cleft. The ligands are represented as yellow sticks and the electron density $2F_o - F_c$ maps are contoured at 1σ . The residues involved in ligand binding are labeled. The schematic representation (*right*) shows the aromatic residues defining the hydrophobic platform in red.

fiber diffraction analysis of xylan (44) and characterized by the internal hydrogen bonds from O3*n* to O5 *n* + 1, which is especially evident in subsites S1 to S3. This conformation is stabilized by Tyr-217 and Trp-308, both positioning the central xylose units, whereas Tyr-303, at the other face of the platform, is stacking to the glycosidic bond linking both xyloses at subsites S2 and S3.

Apart from the hydrophobic stacking interactions, there are relatively few polar contacts involved in ligand binding, as shown in Fig. 5*a*. The xylose unit located at subsite S1 only makes a single direct polar interaction through its O3 hydroxyl with the Arg-189 peptide carbonyl. Also, the sugar O2 atom makes some packing polar interactions in molecules A and C, but not in molecule B, in which the xylose ring occupying this subsite is not clearly seen in the electron density map. Consequently, it is apparent that the xylose observed in molecules A and C may be partially fixed by packing interactions and not by a tight interaction within this subsite. On the contrary, subsite S2 is mainly defined by stacking to Tyr-217 but is further stabilized by the O5 hydrogen bond to Tyr-268-OH and Arg-189-NH₂. Similarly, the xylose unit at subsite S3 is positioned by stacking to Trp-308, but its binding is further enhanced by tight interactions of the O2 hydroxyl to Tyr-303 and Arg-189 side chains, and a bifurcated hydrogen bond of Glu-305 carboxylate to both the O2 and O3 hydroxyls. Finally, the xylose at subsite S4 is only interacting through its O2 hydroxyl with the peptide CO

of the Tyr-268 main chain. Again, some packing interactions are also observed in molecule A, which creates a different environment possibly producing the diverse conformations observed at this subsite and commented below.

Soaking with xylotriose led to the sugar occupying subsites S2 to S4 (Fig. 5*b*). The interaction pattern is conserved with that described for xylotetraose. However, only molecule A shows the oligosaccharide at full occupancy in all three subsites, as molecules B and C present poor density at subsite S4, which may be attributed to the different packing environment described above. These results led to the conclusion that positions S2 and S3 are tightly held at the binding cleft, whereas xyloses at subsites S1 and S4 are more loosely bound. Also, an inspection to the complex reveals that only O2 and O3 hydroxyls are engaged in polar interactions with the protein at subsite S3, whereas all other xylose units situate these hydroxyls completely exposed to the solvent. Consequently, and with the sole exception of subsite S3, the CBM22-2 binding cleft appears able to allocate the full repertoire of side chains evident in xylans.

Moreover, soaking experiments with the tetrasaccharide Glc-4Glc-4Glc-3Glc yielded rather disordered crystals in which only the central portion, including two β -1,4-linked glucose units, could be modeled in molecule A (Fig. 5*c*). The two glucose rings are located at subsites S2 and S3 as described previously, but their positions are shifted along the hydrophobic platform with respect to that observed in the xylooligosaccha-

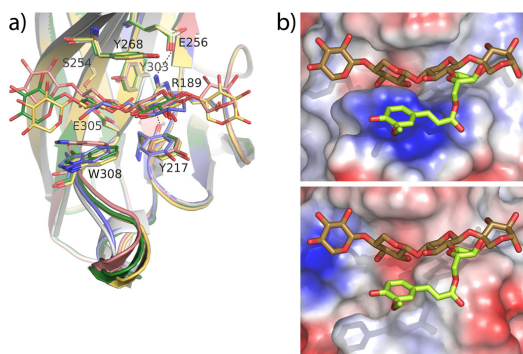


FIGURE 6. **CBM22-2-binding site architecture.** *a*, superposition of the CBM22-2 complexes onto the free CBM22-2 (white). Two positions of xyl4 (pink and yellow), xyl3 (green), and Glc-4Glc-4Glc-3Glc (blue) are shown. A face of the binding cleft is fixed through polar interactions of the involved residues, although the opposite face is defined by a plastic hydrophobic platform able to fit the substrate. *b*, top, CBM22-2 binding cleft showing the experimental position of xyl4 and a modeled arabinose-ferulate moiety (lime) decorating O₂ at subsite S2. Bottom, CBM22-1 binding cleft showing a putative position of the same ligand, inferred from structural superimposition of both domains.

rides. First, the glucose occupying subsite S2 moves away from Arg-189 to allocate its C6-OH hydroxyl group while losing the polar interactions made by the cyclic O5 from xylose. In this way, Tyr-217 is stacking against the C3-C4-C5 portion of the glucose ring. Second, the glucose at subsite S3 is now stacking against Tyr-268, at the other face of the platform, but maintains hydrogen bonds through its O2 and O3 hydroxyls to Arg-189, Tyr-303, and Glu-305, although with a different interaction pattern (Fig. 5c). Nevertheless, and because of the poor density maps, we can not exclude the presence of other conformations bound at the cleft.

Superposition of the different ligand complexes allows depiction of the molecular mechanism of substrate recognition by CBM22-2. As shown on Fig. 6a, one face of the cleft is formed by a series of polar residues that form a network of hydrogen bonds, which maintains a tight arrangement of the residues essential for binding. Thus, the Arg-189 side chain is fixed by hydrogen links to Tyr-217 and Glu-256 through its NH1 and NH2, whereas the position of Glu305 is fixed by interaction with Ser-254. Moreover, the stacking pattern interaction through Ser-254, Tyr-303, and Arg-189 contributes to fixing this conformation. On the contrary, the other face of the cleft is defined by a plastic hydrophobic platform formed by Tyr-217 and Trp-308. In fact, mobility of Trp-308 seems to be facilitated by flexibility of the loop 308–313, whose conformation has been observed to be variable in the crystals. Therefore, the inserted loop, including Trp-308, is an essential determinant bringing specificity to CBM22-2. Furthermore, the plasticity of the hydrophobic platform base, and the looser binding observed at subsites S1 and S4, might be the structural basis explaining why CBM22-2 is able to accommodate mixed glucans.

CBM22-2 Domain Presents Novel Xylan Binding Determinants within the CBM22 Family—The 23% sequence identity shared by CBM22-1 and CBM22-2 is similar to the values obtained when comparing each domain to other CBM22 members, with sequence identities ranging from 15 to 30%. Yet

topology is rather conserved, and both domains show root mean square deviations of only 1.8–1.9 Å (over 145–149 Cα atoms) when superimposed on the two other structurally known, CBM22-1/CBM22-2 from *RthXyn10B* (PDB codes 1DYO and 2W5F). This prominent topology conservation is not related to specificity as *RthCBM22-1*, like many CBM22 members, does not contain the five residues identified as essential for ligand binding in *RthCBM22-2*, thus explaining the inability of some CBM22 modules to bind xylan (21). In this work, we have identified Trp-308, as an additional determinant directly involved in CBM22-2 binding to ligands. A Blast search shows that this Trp is found only in a low number of the closest CBM22-2 homologues (less than one-fourth), mostly present in modular GH10-containing enzymes from *Paenibacillus* or *Clostridium* species that also contain the (W/Y)YY motif. The finding that this Trp is missing in CBM22-1 and *RthCBM22-2* sequences, both known to bind xylan, suggests that subtle differences in sequence may allow a fine-tuning of substrate targeting among xylan-binding CBM22 domains.

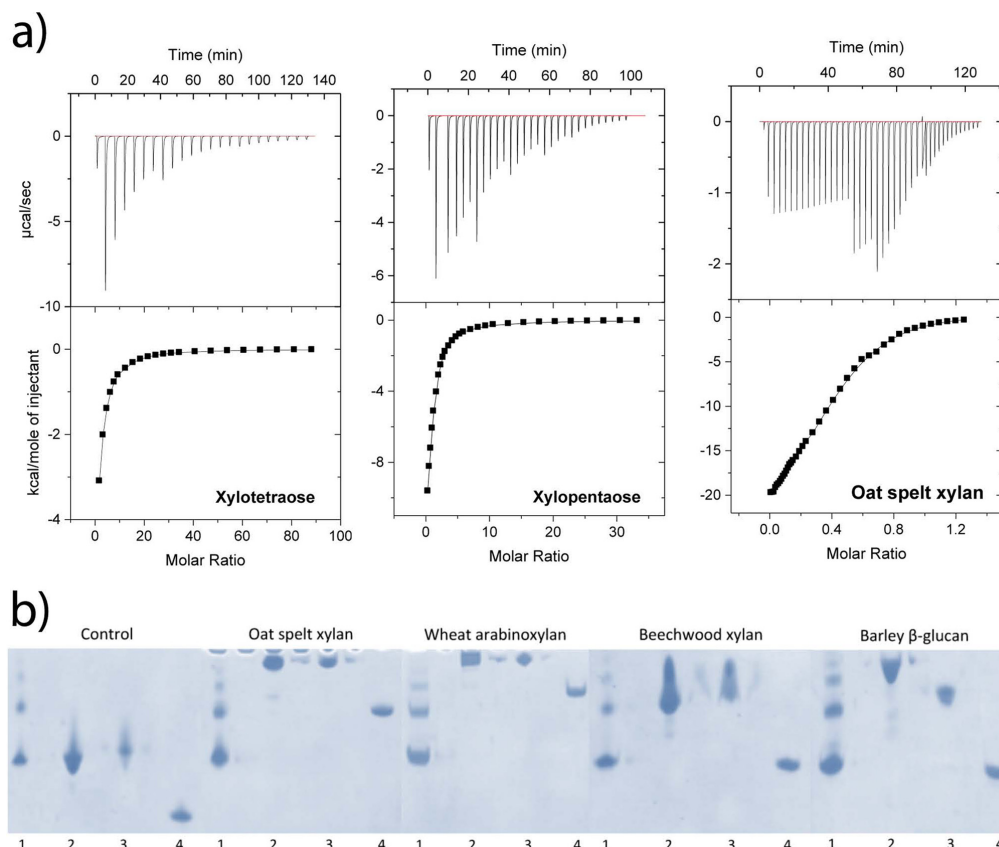
Fig. 6b shows the profile of the CBM22-2 binding cleft, as compared with its partner CBM22-1. As shown in the figure, the flexible loop comprising residues 308–313 protrudes from the CBM22-2 binding crevice in a way that, apart from fitting the xylose unit at subsite S3, it also contours a platform that could help in creating a site to allocate the xylan substitutions of the xylose bound at subsite S2. As in CBM22-1, this loop is much shorter in most CBM22s that, consequently, leave room to tolerate large xylan substitutions at subsite S2. The special feature of the CBM22-2-type domains is that they might be able to recognize, or even select, a particular substitution of xylan. This ability is illustrated by a putative arabinose-ferulate decoration modeled at O₂, represented in Fig. 6b, which seems to be accommodated nicely onto the binding platform. Thus, the loop insertion and the presence of an aromatic residue at the position equivalent to Trp-308 might be a common trait in some CBM22 members accounting for preferential binding to a particular region of the heterogeneous multivalent polymer, which has yet to be discovered.

Binding Affinity of Xyn10C-XBD and Its Isolated Modules—The binding ability of Xyn10C-XBD for short-chain oligosaccharides was evaluated by ITC. The binding isotherms for each ligand were fitted by nonlinear regression assuming two independent sites (one per domain) in the tandem (Table 3 and Fig. 7a). The stoichiometry of binding was experimentally confirmed for *K* values on the order of 10⁴ M^{−1} or higher. Both domains bind xylotriose with similar affinity, but one of them recognizes oligosaccharides with four or more xylose units with higher affinity and more favorable enthalpy. Moreover, the binding of xylopentaose or xylohexose was clearly favored over that of xylotetraose. The presence of four subsites at the binding cleft of CBM22-2 can account for the stronger interaction of xylotetraose, further enhanced by the extra contribution of Trp-308 to the complex stabilization when compared with CBM22-1. However, the higher affinity of Xyn10C-XBD against xylopentaose or xylohexose with respect to xylotetraose might be attributed to a higher conservation of the preferred 3-fold helical conformation in the longer oligosaccharides, resulting in higher binding efficiencies, as further subsites

Structure and Function of Xyn10C

TABLE 3
Thermodynamic parameters of Xyn10C-XBD to different ligands at 25 °C

| Ligand | | $K_1 \times 10^{-3}$ | $-\Delta H_1$ | $-T\Delta S_1$ | $K_2 \times 10^{-3}$ | $-\Delta H_2$ | $-T\Delta S_2$ |
|------------|---|------------------------|---------------|----------------|------------------------|---------------|----------------|
| | | [iscap]m ⁻¹ | kcal/mol | kcal/mol | [iscap]m ⁻¹ | kcal/mol | kcal/mol |
| Xyn10C-XBD | Xylotriose ^a | 1.26 ± 0.01 | 7.82 ± 0.04 | 3.60 ± 0.04 | | | |
| | Xylotetraose | 7.8 ± 0.4 | 15.2 ± 0.9 | 9.9 ± 0.5 | 2.1 ± 0.2 | 7.5 ± 0.9 | 2.98 ± 0.8 |
| | Xylopentaose | 31 ± 9 | 16 ± 2 | 10 ± 2 | 2 ± 1 | 8 ± 1 | 3.5 ± 0.7 |
| | Xylohexaose | 36 ± 6 | 11.5 ± 0.8 | 5.31 ± 0.7 | 5.0 ± 0.4 | 8.6 ± 0.8 | 3.6 ± 0.8 |
| | Xylan oat spelt (0.381 mM) ^b | 64 ± 7 | 5.3 ± 0.6 | -1.2 ± 0.5 | 12 ± 1 | 24.8 ± 0.5 | 19.3 ± 0.4 |
| | G3G4G4G ^c | ~0.2 ^c | ~5.9 | 2.6 | | | |
| | G4G4G3G ^c | ~0.2 | ~5.0 | 1.9 | | | |
| CBM22-1 | Xylotetraose | 3.6 ± 0.2 | 17.42 ± 0.04 | 12.58 ± 0.01 | | | |
| | Xylotetraose | 1.6 ± 0.2 | 10.07 ± 0.08 | 5.71 ± 0.01 | | | |

^a Data were fit to the onset of sites model with a 1:1 stoichiometry per CBM22 domain.^b Millimolar concentration of the two types of sites in the polysaccharide that gives a value of binding sites per CBM22 domain equal to 1.^c Reported values are estimations of the thermodynamic parameters due to the incomplete (~50–60%) saturation of sites reached in ITC with the mixed oligosaccharides.**FIGURE 7. Binding specificity of Xyn10C-XBD and its isolated domains.** *a*, representative binding of different carbohydrates to Xyn10C-XBD by ITC. The ligands in the syringe (20 mM oligosaccharides, 4 mg/ml polysaccharide) were titrated into Xyn10C-XBD loaded in the cell (46–60 μM). Measurements were performed at 25 °C in 50 mM sodium hydrogen/dihydrogen phosphate, pH 7. The upper panels show the raw data upon injection of 1, 5 × 5, 8 × 10, and 8 × 20 μl for xylotetraose; 1 × 2, 5 × 3, 5 × 5, 4 × 10, and 9 × 20 μl for xylopentaose; and 1, 15 × 5, 4 × 10, and 21 × 20 μl for oat spelt arabinoxylan. The bottom panels show the integrated areas (symbols) obtained from respective raw data and their theoretical fit (continuous lines) using the two-sets of sites binding-model of the Origin ITC software with the best fitting thermodynamic parameters summarized in Table 3. *b*, nondenaturing polyacrylamide gels containing no ligand (control) or soluble xylans. Lane 1, BSA; lane 2, Xyn10C-XBD; lane 3, CBM22-1; lane 4, CBM22-2.

were not evident from the CBM22-2 complexes. Interestingly, titration experiments also showed that the CBM22 tandem of Xyn10C-XBD binds xylooligosaccharides with higher affinity

than mixed glucans. This reduced affinity for glucans is consistent with the disorder observed in the crystals by soaking experiments with the tetrasaccharide Glc-4Glc-4Glc-3Glc, and it

Structure and Function of Xyn10C

probably reflects a statistical occupancy of different suboptimal binding modes. Binding against a polymeric substrate, the oat spelt arabinoxylan was also assessed. The titration curve was best fitted by assuming two types of sites whose affinities for the polysaccharide exceeded those of xylopentaose or xylohexose, but the ratios between the binding constants of both sites were comparable. To further characterize the individual Xyn10C-XBD-binding sites, binding of each independent CBM22 domain was tested against xylotetraose by ITC. Affinities were similar and around 2–5 times lower than for the tandem highest affinity site, excluding remarkable cooperative effects between both binding sites within the tandem (Table 3).

Binding ability to xylans and barley of each isolated CBM22 domain was qualitatively evaluated using affinity gel electrophoresis (Fig. 7b). Migration of both CBM22-1 and CBM22-2 was markedly retarded with respect to migration in gels without the substrates, especially when arabinoxylans are tested; in this case, larger retardation to the highly substituted wheat arabinoxylan was apparent in CBM22-2. Moreover, the binding observed in gels for each domain was similar and equivalent to that observed for the tandem when tested against beechwood glucuronoxylan. On the contrary, Xyn10C-XBD was more retarded in barley β -glucan than each isolated CBM22.

Finally, a fact worth mentioning is that CBM22-1 migrates in absence/presence of substrates similarly to Xyn10C-XBD and is distinctly compared with that observed for its partner CBM22-2. It is tentatively suggested that the reason for this observation might be attributed to the common association state suggested by the crystal analysis presented here.

Catalytic Domain May Bind Highly Decorated Substrates—To further elucidate the catalytic functionality of Xyn10C, a construct containing its GH10 domain (residues 340–691) has been solved at 1.76 Å resolution. The structure of the catalytic domain corresponds to the expected topology of a single $(\beta/\alpha)_8$ barrel, which is common in the GH10 family. The barrel is closely packed on the face corresponding to the short loops linking the different β/α repeats (Fig. 8a), whereas the long segments at the C terminus of the β -strands (L1 to L8) present additional elements of secondary structure and form the active site. The catalytic residues are located at the end of β_4 (the general acid/base Glu-475) and β_7 (the nucleophile Glu-593) strands, respectively. A special feature of Xyn10C-CD is the 10-residue insertion found in loop L4. Most GH10 members present a short α -helix just prior to α_4 , but an aspartate-rich loop (⁴⁹⁸DVDGDDGDDSD⁵⁰⁷) is inserted between these two helices in Xyn10C-CD, constituting a nonconserved calcium-binding site (Fig. 8a). A structural calcium ion conferring stability to *Cellvibrio japonicus* Xyn10A has been reported (45), but that metal ion is bound by residues from loop L7 (PDB code 1W2P).

As observed in GH10 xylanases, the shape of the α/β -barrel is elliptical, with the long L7 and L8 loops shaping an extended open cleft consistent with its endo-mode of attack and able to accommodate polymeric substrates. The subsite mapping of the binding cleft is illustrated by superimposition of the XynC-CD coordinates to those of the xylanase Xyn10b from *Cellvibrio mixtus* (46) (37% sequence identity), in complex with arabinose-substituted xylo-oligosaccharides (PDB code 1UR2). In this complex, arabinose- α -1,3-xylotriose and xylotriose were

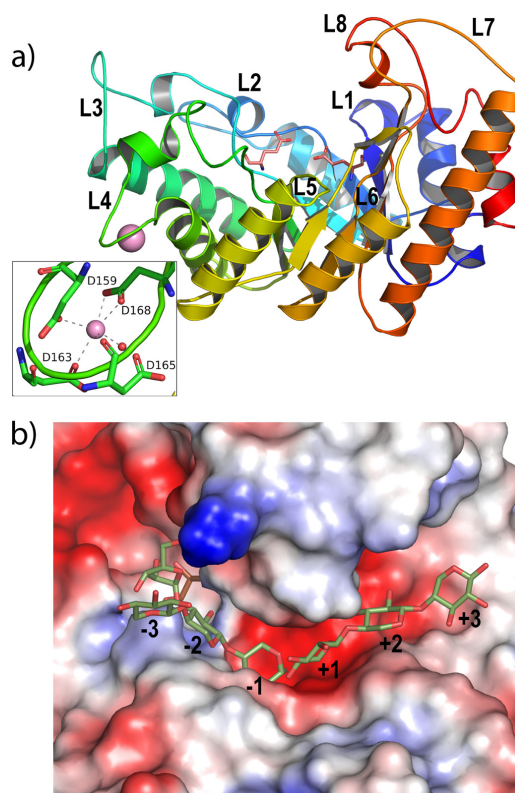


FIGURE 8. Structure of Xyn10C-CD. a, catalytic domain (represented in rainbow sequence code) folds into a $(\beta/\alpha)_8$ barrel common to GH10 family, with the catalytic Glu-136 and Glu-254 being at the end of β_4 and β_7 . The long segments at the end of the β -strands (L1 to L8) are shaping an open active site cleft. An insertion at loop L4 is defining a calcium ion (violet sphere)-binding site, zoomed in the inset. b, mapping of the active site by superimposition of Xyn10C-CD onto *CmXyn10b* (46) complexed with arabinose- α -1,3-xylotriose/xylotriose at the glycone (minus)/aglycone (plus) subsites. A MPD molecule (pink) from the crystallization buffer is occupying subsite -2 in Xyn10C-CD crystals.

observed at the glycone (minus) and aglycone (plus) subsites, respectively, following the nomenclature used for sugar-binding subsites, which defines that the cleavage occurs between subsite -1 and +1 (Fig. 8b). The inspection to the Xyn10C-CD binding cleft reveals that the enzyme may allocate at least six xylose units at the crevice. Previous studies on *C. mixtus* xylanase Xyn10b (46) and other GH10 xylanases (47) bound to decorated xylo-oligosaccharides have shown that xylan substitutions can be accommodated only in selected subsites in which the O2 and O3 of xylose point into the solvent, i.e., -3, +1, and +3, whereas in the case of the -1 and +2 subsites, the hydroxyls are directed into the protein surface precluding the binding of side-chain xylan decorations. In subsite -2, only O3 is solvent-exposed, whereas most of the GH10s have a glutamate at loop L2 that makes a hydrogen bond to O2, which consequently avoids decoration at this position. However, Xyn10C-CD is one of the reported GH10s having the Glu/Gly

Structure and Function of Xyn10C

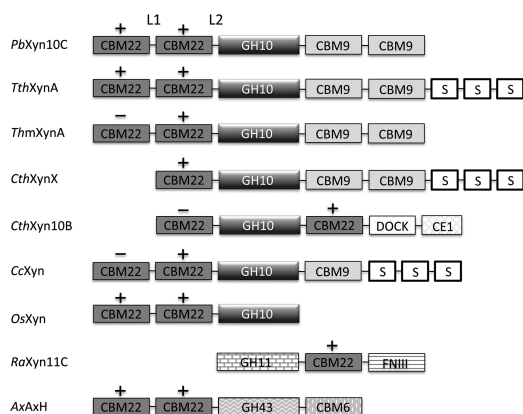


FIGURE 9. Modular structure of representative CBM22 containing enzymes. *P. barcinonensis* xylanase C (PbXyn10C); *Thermoanaerobacterium thermosulfurigenes* xylanase A (TthXynA, Q60046); *Thermotoga maritima* xylanase A (TmXynA, Q60037); *Clostridium thermocellum* xylanase X (CthXynX, P38535); *C. thermocellum* xylanase 10B (CthXyn10B, P51584); *Clostridium cellulolyticum* xylanase (CcXyn, B815C0); *Oryza sativa* xylanase (OsXyn, Q84R74); *Ruminococcus albus* xylanase 11C (RaXyn11C, BAM46898.1); *Amphibacillus xylanus* arabinoxylan arabinofuranohydrolase (AxXynH). S, S-layer-like domain; DOCK, docking domain; CE1, carboxyl esterase 1; FNIII, fibronectin type III. L1, loop linking CBM22-1 to CBM22-2; L2, loop linking CBM22-2 to GH10. +/–, the xylan binding motif (W/Y)YY is present/absent in CBM22 domain.

substitution that provides limited access to O2 (48). In the case of Xyn10C-CD, a short L1 loop leaves additional room that seems able to harbor large substituents at both O2/O3 of subsite –2. A similar broad cavity is observed at subsite +2, which is occupied by several ordered water molecules in the crystal. Consequently and similarly to what it is observed in the N-terminal domain, the catalytic module of Xyn10C seems able to bind highly decorated xyans by accommodation of substitutions in almost all its subsites.

Discussion

Structural studies on representative members of many CBMs and their complexes have provided insight into the underlying mechanism of CBM-ligand recognition and interaction. At present, one of the greatest challenges is to depict the concerted molecular mechanism that displays multimodular biomass-degrading enzymes that present multiple copies of ancillary non-catalytic domains. Homogeneous multimodularity has been related to multivalency and avidity effects, whereas the heterogeneous pattern has been proposed to provide distinct ligand-binding specificities. However, previous studies have revealed that tandem of homologous CBMs may present a more complex functional mechanism (49), thereby suggesting that the apparent functional distinction between the heterogeneous and homogeneous modular arrangement might not be as simple.

The xylanase Xyn10C from *P. barcinonensis* is a 120-kDa modular enzyme that presents the CBM22/GH10/CBM9 architecture found in a subset of large GH10 xylanases, mainly produced by saprophytic microorganisms (Fig. 9). Apart from the group of enzymes found in thermophiles, with a regular domain's composition, a more variable modular architecture is found in some members from *Clostridium* species where, in some

cases, two CBM22s are flanking the catalytic GH10 domains as occurs in CthXyn10B. Interestingly, some of these CBM22s do not present the three-aromatic motif associated with xylan binding, therefore pointing to yet unknown functions. In a few cases, the CBM22 domain is present in bacterial enzymes involved in xylan degradation bearing catalytic domains from families other than GH10, as GH11 or GH43. Furthermore, a number of plant xylanases encompass one to four CBM22 repeats preceding their catalytic GH10. This diversity reflects the existence of common recognition mechanisms that must be combined with particular determinants specific to each function.

We report here the three-dimensional structure of the Xyn10C N-terminal region, containing the CBM22-1–CBM22-2 tandem, which was shown to bind xylan, and hence we called it Xyn10C-XBD. This represents the first crystal structure of two contiguous CBM22 domains solved to date. Xyn10C-XBD is folded into two separate CBM22 domains that are linked by a flexible 10-residue segment, which provides an extraordinary plasticity to the domain that consequently exhibits a broad conformational landscape. This feature must potentiate significantly the approach of each CBM moiety to their corresponding targets.

Each CBM22 domain of Xyn10C-XBD presents the expected β -jelly roll topology and both contain the R(W/Y)YYE motif that had previously been identified to be required for xylan binding. However, the soaking analysis of the isolated CBM22-2 crystals with ligands has allowed us to identify the Trp-308 as an additional determinant for binding. This residue is located in an insertion found in a low number of closest CBM22-2 homologues mostly integrated in modular GH10-containing enzymes from *Paenibacillus* or *Clostridium* species that also present the xylan-binding motif. Consequently, the CBM22-2 domain may define a subset of xylan-binding CBM22s directed to precise regions of the polysaccharide. There might be other structural determinants fine-tuning xylan binding functionality within the CBM22 family that have yet to be discovered.

The structural differences found in both Xyn10C-XBD-binding sites are consistent with the ITC experiments that show the presence of two types of sites with different affinity. Consequently, it is most likely that both domains account for a different functionality *in vivo*, explaining the need of domain multiplicity in Xyn10C. Moreover, the affinity gel results show that both the tandem and its isolated domains were noticeably retarded when tested with arabinoxylans. However, it did not show clear differences between the isolated CBM22s, apart from a slight preference of CBM22-2 for highly substituted arabinoxylan. Nevertheless, it has been previously reported that CBMs showing equivalent specificity against purified substrates exhibit significant differences in their recognition of plant cell walls (50). Therefore, the precise and distinct physiological role of CBM22-1 versus CBM22-2 may elude analyses performed with polysaccharides isolated from their natural environment.

The different structural determinants described in this work for each CBM22 moiety of Xyn10C suggest that the tandem might represent a sophisticated delivery strategy of Xyn10C in which CBM22-1 would represent a primary recognition site

directing the enzyme to a particular region or a specific type of cell wall. Then the CBM22-2 domain would recognize specific regions of the cell wall that are going to be degraded by the GH10 domain, enhancing its catalytic efficiency. The great flexibility of the CBM22-1–CBM22-2 linker could enable the CBM22-2/GH10 domains to access a large extension of the cell wall providing a greater chance of target substrate capturing. In agreement with this, the shorter link attaching the CBM22 tandem to GH10 (³³²TEAIAIE³³⁸) is indicative of a more restricted conformational freedom at this segment and a more concerted action of the CBM22-2/GH10 domains, both of which are apparently able to target highly decorated glucuronoarabinoxylans. These polysaccharides are major components of the primary cell walls in cereals and, therefore, they probably are the most abundant substrate found in the natural habitat of Xyn10C.

As more structural studies on CBMs become available, it is more evident that the information obtained from the isolated modules must necessarily be placed in the context of the whole enzyme. In particular, it has been suggested that the size and chemical nature of linkers connecting different CBMs and/or their associated catalytic domains may reveal functional features on a putative integrated role between modules (2). In this sense, it is interesting to note that the Xyn10C homologues from *Bacillus* containing the same CBM22–CBM22–GH10–CBM9–CBM9 architecture present a conserved configuration in its N-terminal region with loops following the pattern L1 > L2 (Fig. 9), whereas, their homologues from thermophiles present L1 < L2. Therefore, the integrated role of CBM22-2/GH10 domains proposed for Xyn10C may be a trait of its homologues from mesophilic bacteria but not from thermophiles. However, the occurrence of the common additional C-terminal cellulose-directed CBM9 domain would confer to all of them the additional advantage of allowing the enzymes to remain in close contact with the cell wall materials, as CBM9s can diffuse over the surface of the cellulose microfibrils.

In conclusion, the analysis of Xyn10C and its CBM22-1–CBM22-2 tandem domains presented here provides novel features that may help to understand the intricate molecular mechanisms displayed by enzymes containing multiples copies of CBMs. Indeed, more work is necessary to have a clear picture of the sophisticated strategy used by nature to tailor specificity against the highly complex and heterogeneous polysaccharides forming the plant cell wall. This basic knowledge is crucial to take the full potential of biocatalysts to produce biofuel and more efficient bioprocessing industries.

Acknowledgments—We thank Randy Read and Paul Adams for their valuable help in solving the structure of the Xyn10C–XBD crystals. We thank the Spanish Synchrotron at Barcelona (ALBA, Spain) for assistance at the XALOC beamline. We also thank the staff of the European Synchrotron Radiation Facility at Grenoble (ESRF, France) for providing access and technical assistance at beamlines ID23-1, ID23-2, and ID29. We thank Dr. Douglas Laurents for kindly revising the manuscript.

References

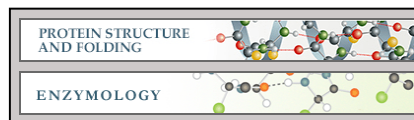
- Gilbert, H. J. (2010) The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiol.* **153**, 444–455
- Abbott, D. W., and van Bueren, A. L. (2014) Using structure to inform carbohydrate binding module function. *Curr. Opin. Struct. Biol.* **28**, 32–40
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495
- Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* **382**, 769–781
- Gilbert, H. J., Knox, J. P., and Boraston, A. B. (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Curr. Opin. Struct. Biol.* **23**, 669–677
- Ficko-Blean, E., and Boraston, A. B. (2012) Insights into the recognition of the human glycome by microbial carbohydrate-binding modules. *Curr. Opin. Struct. Biol.* **22**, 570–577
- Hervé, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J., and Knox, J. P. (2010) Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15293–15298
- Fontes, C. M., and Gilbert, H. J. (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem.* **79**, 655–681
- Vazana, Y., Morais, S., Barak, Y., Lamed, R., and Bayer, E. A. (2012) Designer cellulosomes for enhanced hydrolysis of cellulosic substrates. *Methods Enzymol.* **510**, 429–452
- Shoseyov, O., Shani, Z., and Levy, I. (2006) Carbohydrate binding modules: biochemical properties and novel applications. *Microbiol. Mol. Biol. Rev.* **70**, 283–295
- Vaaje-Kolstad, G., Horn, S. J., van Aalten, D. M., Synstad, B., and Eijsink, V. G. (2005) The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. *J. Biol. Chem.* **280**, 28492–28497
- Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sørlie, M., and Eijsink, V. G. (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* **330**, 219–222
- Montanier, C., van Bueren, A. L., Dumon, C., Flint, J. E., Correia, M. A., Prates, J. A., Firbank, S. J., Lewis, R. J., Grondin, G. G., Ghinet, M. G., Gloster, T. M., Herve, C., Knox, J. P., Talbot, B. G., Turkenburg, J. P., et al. (2009) Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3065–3070
- Simpson, P. J., Bolam, D. N., Cooper, A., Ciruela, A., Hazlewood, G. P., Gilbert, H. J., and Williamson, M. P. (1999) A family IIb xylan-binding domain has a similar secondary structure to a homologous family IIa cellulose-binding domain but different ligand specificity. *Structure* **7**, 853–864
- Szabo, L., Jamal, S., Xie, H., Charnock, S. J., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2001) Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation. *J. Biol. Chem.* **276**, 49061–49065
- Sainz-Polo, M. A., Valenzuela, S. V., González, B., Pastor, F. I., and Sanz-Aparicio, J. (2014) Structural analysis of glucuronoxylan-specific Xyn30D and its attached CBM35 domain gives insights into the role of modularity in specificity. *J. Biol. Chem.* **289**, 31088–31101
- Fontes, C. M., Hazlewood, G. P., Morag, E., Hall, J., Hirst, B. H., and Gilbert, H. J. (1995) Evidence for a general role for non-catalytic thermostabilizing domains in xylanases from thermophilic bacteria. *Biochem. J.* **307**, 151–158
- Meissner, K., Wassenberg, D., and Liebl, W. (2000) The thermostabilizing domain of the modular xylanase XynA of *Thermotoga maritima* represents a novel type of binding domain with affinity for soluble xylan and mixed-linkage β -1,3/ β -1,4-glucan. *Mol. Microbiol.* **36**, 898–912
- Sunna, A., Gibbs, M. D., and Bergquist, P. L. (2000) The thermostabilizing domain, XynA, of *Caldibacillus cellulosovorans* xylanase is a xylan binding domain. *Biochem. J.* **346**, 583–586
- Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M., Davies, G. J., and Fontes, C. M. (2000) The X6 “thermostabilizing” domains of xylanases are carbohydrate-binding modules: structure and

Structure and Function of Xyn10C

- biochemistry of the *Clostridium thermocellum* X6b domain. *Biochemistry* **39**, 5013–5021
21. Xie, H., Gilbert, H. J., Charnock, S. J., Davies, G. J., Williamson, M. P., Simpson, P. J., Raghothama, S., Fontes, C. M., Dias, F. M., Ferreira, L. M., and Bolam, D. N. (2001) *Clostridium thermocellum* Xyn10B carbohydrate-binding module 22-2: the role of conserved amino acids in ligand binding. *Biochemistry* **40**, 9167–9176
 22. Sánchez, M. M., Fritze, D., Blanco, A., Spröer, C., Tindall, B. J., Schumann, P., Kroppenstedt, R. M., Díaz, P., and Pastor, F. I. (2005) *Paenibacillus barcinonensis* sp. nov., a xylanase-producing bacterium isolated from a rice field in the Ebro River delta. *Int. J. Syst. Evol. Microbiol.* **55**, 935–939
 23. Gallardo, O., Pastor, F. I., Polaina, J., Díaz, P., Yšek, R., Vogel, P., Isorna, P., González, B., and Sanz-Aparicio, J. (2010) Structural insights into the specificity of Xyn10B from *Paenibacillus barcinonensis* and its improved stability by forced protein evolution. *J. Biol. Chem.* **285**, 2721–2733
 24. Valenzuela, S. V., Díaz, P., and Pastor, F. I. (2014) Xyn11E from *Paenibacillus barcinonensis* BP-23: a LppX-chaperone-dependent xylanase with potential for upgrading paper pulps. *Appl. Microbiol. Biotechnol.* **98**, 5949–5957
 25. Blanco, A., Díaz, P., Zueco, J., Parascandola, P., and Javier Pastor, F. I. (1999) A multidomain xylanase from a *Bacillus* sp. with a region homologous to thermostabilizing domains of thermophilic enzymes. *Microbiology* **145**, 2163–2170
 26. Hogg, D., Pell, G., Dupree, P., Goubet, F., Martín-Orúe, S. M., Armand, S., and Gilbert, H. J. (2003) The modular architecture of *Cellvibrio japonicus* mannanases in glycoside hydrolase families 5 and 26 points to differences in their role in mannan degradation. *Biochem. J.* **371**, 1027–1043
 27. Correia, M. A., Abbott, D. W., Gloster, T. M., Fernandes, V. O., Prates, J. A., Montanier, C., Dumon, C., Williamson, M. P., Tunnicliffe, R. B., Liu, Z., Flint, J. E., Davies, G. J., Henriessat, B., Coutinho, P. M., Fontes, C. M., and Gilbert, H. J. (2010) Signature active site architectures illuminate the molecular basis for ligand specificity in family 35 carbohydrate binding module. *Biochemistry* **49**, 6193–6205
 28. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685
 29. Valenzuela, S. V., Díaz, P., and Javier Pastor, F. I. (2010) Recombinant expression of an alkali stable GH10 xylanase from *Paenibacillus barcinonensis*. *J. Agric. Food Chem.* **58**, 4814–4818
 30. Monterroso, B., Sáiz, J. L., García, P., García, J. L., and Menéndez, M. (2008) Insights into the structure-function relationships of pneumococcal cell wall lysozymes, LytC and Cpl-1. *J. Biol. Chem.* **283**, 28618–28628
 31. Sainz-Polo, M. Á., González, B., Pastor, F. I., and Sanz-Aparicio, J. (2015) Crystallization and preliminary x-ray diffraction analysis of the N-terminal domain of *Paenibacillus barcinonensis* xylanase 10C containing the CBM22-1–CBM22-2 tandem. *Acta Crystallogr. F Struct. Biol. Commun.* **71**, 136–140
 32. Batty, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281
 33. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
 34. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242
 35. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
 36. Najmudin, S., Pinheiro, B. A., Prates, J. A., Gilbert, H. J., Romão, M. J., and Fontes, C. M. (2010) Putting an N-terminal end to the *Clostridium thermocellum* xylanase Xyn10B story: crystal structure of the CBM22-1-GH10 modules complexed with xylohexase. *J. Struct. Biol.* **172**, 353–362
 37. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
 38. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
 39. Vagin, A. A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025
 40. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291
 41. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21
 42. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, Version 1.3, DeLano Scientific, San Carlos, CA
 43. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
 44. Atkins, E. D. T. (1992) *Xylan and Xylanases: Progress in Biotechnology*, Vol. 7, pp. 39–50, Elsevier Science Publishers, Amsterdam
 45. Andrews, S. R., Taylor, E. J., Pell, G., Vincent, F., Ducros, V. M., Davies, G. J., Lakey, J. H., and Gilbert, H. J. (2004) The use of forced protein evolution to investigate and improve stability of family 10 xylanases. The production of Ca²⁺-independent stable xylanases. *J. Biol. Chem.* **279**, 54369–54379
 46. Pell, G., Taylor, E. J., Gloster, T. M., Turkenburg, J. P., Fontes, C. M., Ferreira, L. M., Nagy, T., Clark, S. J., Davies, G. J., and Gilbert, H. J. (2004) The mechanisms by which family 10 glycoside hydrolases bind decorated substrates. *J. Biol. Chem.* **279**, 9597–9605
 47. Vardakou, M., Flint, J., Christakopoulos, P., Lewis, R. J., Gilbert, H. J., and Murray, J. W. (2005) A family 10 *Thermoascus aurantiacus* xylanase utilizes arabinose decorations of xylan as significant substrate specificity determinants. *J. Mol. Biol.* **352**, 1060–1067
 48. Pell, G., Szabo, L., Charnock, S. J., Xie, H., Gloster, T. M., Davies, G. J., and Gilbert, H. J. (2004) Structural and biochemical analysis of *Cellvibrio japonicus* xylanase 10C: how variation in substrate-binding cleft influences the catalytic profile of family GH-10 xylanases. *J. Biol. Chem.* **279**, 11777–11788
 49. Abbott, D. W., Erin-López, J. M., and Boraston, A. B. (2008) Insight into ligand diversity and novel biological roles for family 32 carbohydrate-binding modules. *Mol. Biol. Evol.* **25**, 155–167
 50. McCartney, L., Blake, A. W., Flint, J., Bolam, D. N., Boraston, A. B., Gilbert, H. J., and Knox, J. P. (2006) Differential recognition of plant cell walls by microbial xylan-specific carbohydrate-binding modules. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4765–4770



Protein Structure and Folding:
**Exploring Multimodularity in Plant Cell
Wall Deconstruction: STRUCTURAL AND
FUNCTIONAL ANALYSIS OF Xyn10C C
ONTAINING THE CBM22-1–CBM22-2
TANDEM**



M. Angela Sainz-Polo, Beatriz González,
Margarita Menéndez, F. I. Javier Pastor and
Julia Sanz-Aparicio
J. Biol. Chem. 2015, 290:17116–17130.
doi: 10.1074/jbc.M115.659300 originally published online May 22, 2015

Access the most updated version of this article at doi: [10.1074/jbc.M115.659300](https://doi.org/10.1074/jbc.M115.659300)

Find articles, minireviews, Reflections and Classics on similar topics on the [JBC Affinity Sites](#).

Alerts:
• [When this article is cited](#)
• [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 49 references, 22 of which can be accessed free at
<http://www.jbc.org/content/290/28/17116.full.html#ref-list-1>

4.2.3. Resultados

En este trabajo se presenta el análisis estructural de Xyn10C, una xilanas modular con una composición de dominios CBM22/GH10/CBM9. Se ha resuelto la estructura tridimensional del dominio N-terminal que contiene el tándem CBM22-1-CBM22-2, denominado Xyn10C-XBD por su capacidad de unir xilanos, de estos módulos aislados con ligandos y del dominio catalítico Xyn10C-CD. Hay que destacar que se trata de la primera estructura reportada que contiene dos módulos CBM22 contiguos en tándem.

El análisis cristalográfico muestra que Xyn10C-XBD se pliega en dos módulos CBM22 separados por un fragmento conector de 10 aminoácidos con una gran movilidad. Esto le confiere una gran flexibilidad y, con ello, una amplia variedad conformacional, lo cuál se observa en el cristal que presenta seis moléculas independientes en la unidad asimétrica con distinta orientación de dominios. Esta característica debe potenciar de forma significativa la orientación de cada CBM con respecto a su sitio de unión. Por otra parte, las moléculas están formando trímeros asimétricos, mediante una asociación regular de los módulos CBM22-1.

Respecto a los experimentos de inmersión realizados con ligandos, únicamente los cristales de CBM22-2 proporcionaron complejos adecuados para investigar su especificidad de unión. En relación con esto es de destacar que el análisis de la estructura del módulo CBM22-1 muestra un patrón de asociación en trímeros idéntico al encontrado en los cristales de Xyn10C-XBD. Además, el sitio de unión al sustrato está localizado cerca de la interfase entre subunidades en el trímero y ocupado por el extremo N-terminal de la subunidad adyacente. Por consiguiente, la difusión de los ligandos en el sitio de unión probablemente causaría una perturbación que podría destruir la integridad del cristal. Esta podría ser la razón por la cual los experimentos de inmersión de los cristales de CBM22-1 con ligandos no progresaron.

Cada CBM22 tiene un plegamiento de sandwich β tipo lectina. Una de las láminas β forma un surco extendido que contiene los tres residuos aromáticos del motivo de unión al sustrato Trp55/Tyr217, Tyr104/Tyr268 y Tyr136/Tyr303 (para CBM22-1/CBM22-2 respectivamente), así como dos residuos conservados: una arginina (Arg27/Arg189) y un glutamato (Glu138/Glu305), que se ha demostrado que son esenciales para el reconocimiento de xilano en esta familia. Al comparar ambos módulos CBM22, y aunque sus secuencias son sólo 39 % similares (23 % idénticas), se puede observar que la topología está bastante conservada y las principales diferencias se localizan en algunos bucles que unen las dos láminas β . Las diferencias en estos lazos alteran radicalmente la afinidad por el sustrato, presentando CBM22-2 una hendidura mucho más profunda. Hay que destacar un bucle presente en este módulo, el cual contiene un residuo aromático, Trp308, que está expuesto al solvente y amplía la plataforma hidrofóbica generando subsitios adicionales de unión no observados previamente en la familia CBM22.

Los ensayos de unión a polisacáridos insolubles desvelaron que sólo el dominio C-terminal tiene la capacidad de unir celulosa mientras que Xyn10C-XBD y Xyn10C-CD tienen afinidad por xilano. Además, los estudios de afinidad por xilanos solubles mediante geles de afinidad (geles de retardo) indicaron que en ningún caso se unían a hidroxietilcelulosa. Así, el mayor retardo se observó frente a arabinoxilanos, presentando CBM22-2 un efecto más notable si se prueba frente a arabinoxilanos muy sustituidos, como en el caso del xilano de trigo. Por su parte, Xyn10C-XBD muestra más retardo en β -glucano de cebada que cada CBM22 aislado, lo que sugiere un efecto aditivo frente a ese sustrato en el tándem. Es interesante resaltar que CBM22-1 migra en ausencia/presencia de sustratos de manera similar a Xyn10C-XBD y de forma diferente a CBM22-2. Esto podría estar relacionado con el estado de asociación común que ambos presentan y que sugiere el análisis cristalográfico.

Las diferencias estructurales y de afinidad que se encuentran en ambos módulos de Xyn10C-XBD son consistentes con los experimentos de calorimetría, los cuales muestran dos sitios de distinta afinidad en el tándem.

Finalmente, el análisis de la estructura cristalográfica del dominio catalítico Xyn10C-CD revela que éste parece capaz de unir xilanos altamente decorados, pudiendo alojar sustituciones en casi todos sus subsitios del centro activo. Por tanto, puede decirse que su especificidad es coherente con la observada en el módulo CBM22-2.

4.2.4. Discusión

El presente trabajo es un claro ejemplo de cómo el estudio estructural de los CBM y sus complejos puede aportar valiosa información acerca del mecanismo de interacción con sus respectivos sustratos. La comprensión del modo de acción de enzimas degradadoras de biomasa es un tema de gran relevancia biotecnológica. En este contexto, el estudio de sistemas modulares con múltiples copias de módulos auxiliares no catalíticos es un aspecto clave para entender las bases moleculares del mecanismo catalítico y de la regulación de la especificidad de las enzimas implicadas.

Los módulos de la familia CBM22 están presentes principalmente en enzimas involucradas en la degradación de xilano acompañando a GH de la familia 10, aunque en ocasiones también forman parte de enzimas de las familias GH11 y GH43. Esta diversidad refleja la existencia de un mecanismo de reconocimiento común que debe combinarse con determinantes específicos para cada función. Además, la naturaleza química del segmento conector entre los distintos módulos debe ser un determinante esencial modulando o limitando la orientación de cada CBM con respecto a sus sustratos específicos.

El patrón de asociación en trímeros que presentan las estructuras cristalográficas de Xyn10C-XBD y CBM22-1 podría indicar la existencia de

oligómeros en Xyn10C. Apoya esta hipótesis la observación en Xyn10C-XBD de un trímero similar al encontrado en CBM22-1 en ausencia de CBM22-2, el cual forma parte de la interfase entre las subunidades de Xyn10C-XBD. Esto sugiere que la tendencia de asociación trimérica observada en ambos cristales podría ser un rasgo característico del módulo CBM22-1 y, por tanto, de la enzima completa. Además, los geles de retardo aparentemente apoyan esta teoría, ya que Xyn10C-XBD y CBM22-1 migran de una manera similar y distinta a CBM22-2. Sin embargo, no hemos conseguido detectar la presencia de especies triméricas mediante el análisis por UC de muestras de Xyn10C o CBM22-1 aislado. Por otra parte, el hecho de que este tipo de asociación impida el acceso de ligandos extendidos al sitio de unión, sugiere que un posible trímero de CBM22-1 pudiera presentar afinidad por cadenas de xilosas terminales que se encuentran accesibles, por ejemplo, en sustratos en degradación. La importancia fisiológica del trímero no es evidente y se necesita reunir más evidencias para averiguar si las formas oligoméricas activas detectadas en el zimograma tienen relación con esta característica.

Un aspecto importante de CBM22-2 es el residuo Trp308 que ha resultado ser un determinante clave en la unión de sustrato. Este residuo no está conservado y se encuentra sólo en algunos homólogos de CBM22-2, los cuales tienen el motivo de unión a xilano y son enzimas modulares GH10 de *Paenibacillus* o *Clostridium*. Este triptófano se encuentra en una inserción de diez residuos que presenta una cierta movilidad. Esta plasticidad podría ser la base estructural para explicar por qué CBM22-2 puede albergar también glucanos mixtos. Además, y con la única excepción del subsitio S3, el surco de unión en CBM22-2 parece ser capaz de poder hospedar el repertorio completo de ramificaciones comunes en xilanos. La especial topología del sitio de unión de CBM22-2, que muestra una plataforma con subsitios adicionales a los descritos en la familia CBM22, nos hace pensar que CBM22-2 representa una nueva subfamilia dentro del grupo que une xilano, con una preferencia de unión a zonas específicas del polisacárido aún por descubrir.

Hay que tener presente que los análisis de afinidad se han realizado siempre frente a sustratos aislados de su entorno natural. Por tanto, es difícil extrapolar de manera exacta el papel fisiológico de CBM22-1 y CBM22-2. Sin embargo, todos los resultados apuntan a que ambos módulos presentan características distintas y que por tanto deben jugar diferente papel en la actividad de Xyn10C, apoyando así la idea de que la asociación de módulos homólogos no implica meramente multiplicidad o multivalencia en la unión al sustrato.

Los diferentes determinantes estructurales que se han descrito en este trabajo para cada dominio CBM22, y su análisis dentro del contexto de la enzima completa Xyn10C, sugieren que el tándem podría representar una sofisticada estrategia de unión al sustrato, en la cual CBM22-1 podría orientar la enzima a un primer sitio de reconocimiento, dirigiéndola a un sitio concreto de la pared celular. Después, CBM22-2 reconocería regiones específicas de ésta que han de ser degradadas por el dominio Xyn10C-CD, mejorando así su eficiencia

catalítica. La gran flexibilidad del conector dentro del tándem podría permitir que el conjunto CBM22-2-GH10 pueda acceder a una gran parte de la pared celular, proporcionando una mayor posibilidad de captura de su sustrato. El segmento conector CBM22-2-GH10 de longitud más pequeña, proporciona una menor flexibilidad, posibilitando quizás que CBM22-2 y Xyn10C-CD actúen de forma concertada, siendo ambos capaces de unir GAXs muy ramificados. Estos sustratos son los mayoritarios en paredes primarias de cereales y, por tanto, es probable que sea el sustrato más abundante encontrado en el hábitat natural de Xyn10C. He hecho de que CBM22-2 sea también capaz de unir glucanos mixtos β -(1,3-1,4) podría proporcionar una ventaja adicional, al ser este el componente mayoritario de las paredes en endospermas de cereales. Finalmente, la presencia del tándem de CBM9 (Xyn10C-CBD) capaz de unir celulosa en el dominio C-terminal, permitiría a la enzima permanecer en estrecho contacto con las microfibrillas de la pared celular vegetal.

En conclusión, el análisis de Xyn10C y su dominio CBM22-1-CBM22-2 presentado en este artículo, proporciona aspectos importantes y novedosos que pueden ayudar en la comprensión de los mecanismos moleculares de actuación de enzimas que poseen múltiples copias de CBM. Para tener un conocimiento claro de esta sofisticada estrategia de especificidad dentro de la red de polisacáridos de la pared celular vegetal, es necesario seguir investigando este tipo de dominios. Este conocimiento básico es crucial para aplicar todo el potencial biocatalizador de estas enzimas y así poder producir biocombustibles de forma más eficiente en las industrias.

4.2.5. Conclusiones: Artículos 1 y 2

Las conclusiones de los trabajos “Artículo 1” y “Artículo 2” se recogen a continuación:

1. Se ha diseñado un protocolo de expresión y purificación del dominio N-terminal de Xyn10C conteniendo el tándem CBM22-1-CBM22-2, que permite obtener muestra de pureza adecuada.
2. Se han obtenido las condiciones de cristalización óptimas para conseguir cristales de calidad suficiente para los experimentos de difracción. Se ha determinado el grupo espacial, $P2_1$, así como sus parámetros de celdilla.
3. El dominio N-terminal de Xyn10C es capaz de unir xilano, por lo que se le ha denominado Xyn10C-XBD.
4. La estructura de Xyn10C-XBD ha sido resuelta mediante reemplazo molecular a 2,4 Å de resolución. La cadena polipeptídica se pliega en dos módulos, cada uno de ellos con estructura de sandwich β tipo lectina.
5. Xyn10C-XBD muestra en el cristal una asociación oligomérica en forma de trímero, del mismo modo que el módulo CBM22-1 aislado.

6. Los dos módulos CBM22-1 y CBM22-2 están unidos en Xyn10C de forma muy flexible gracias a la gran movilidad de su fragmento conector de 10 aminoácidos.
7. Se ha resuelto la estructura cristalográfica de CBM22-1 aislado, con un grupo espacial $H3$. No se han podido obtener cristales de sus complejos, lo que puede deberse a la interacción de residuos del sitio de unión a ligandos con el extremo N-terminal de una subunidad adyacente.
8. Se ha resuelto la estructura tridimensional de CBM22-2 a 1,75 Å, con grupo espacial $P3_2$. Se han obtenido complejos por inmersión de los cristales en xilotetraosa, xilotriosa y el glucano mixto Glc-4Glc-4Glc-3Glc.
9. El sitio de unión de CBM22-2 presenta una topología de surco profundo, cuya base está formada por una plataforma hidrofóbica que presenta cierta plasticidad. Esta plataforma está constituida por una inserción en el bucle 303-308 que define sitios adicionales de unión a los descritos en la familia CBM22.
10. Se ha identificado en CBM22-2 el residuo Trp308, no conservado, como un determinante clave de la especificidad directamente implicado en la unión al sustrato. En base a ello, CBM22-2 define una nueva subfamilia dentro de los módulos CBM22 que unen xilano dirigidos a zonas concretas del polímero.
11. Los ensayos de ITC de Xyn10C-XBD indican que está favorecida la unión de cadenas mayores de cuatro unidades de azúcar así como la unión de XOS frente a la de glucanos mixtos. Además, se observan dos sitios de distinta afinidad tanto frente a oligosacáridos como frente a xilano.
12. Los estudios de calorimetría de CBM22-1 y CBM22-2 aislados mostraron similar afinidad frente a oligosacáridos y comparable al valor obtenido en el sitio con menor afinidad del tándem. Esto puede explicarse considerando fenómenos de cooperatividad.
13. El retardo observado en los geles de afinidad de cada módulo, CBM22-1 y CBM22-2, es similar y equivalente a lo observado en el tándem cuando se investiga frente a xilano de haya (GX). Por el contrario, Xyn10C-XBD es más retenido en β -glucano de cebada (GAX) que cada CBM22 aislado, lo que sugiere un efecto aditivo frente a ese sustrato en el tándem.
14. CBM22-1 migra en ausencia/presencia de sustratos de manera similar a Xyn10C-XBD y de forma diferente a su compañero CBM22-2. Esto sugiere un estado de asociación común en ambos, tal como sugiere el análisis cristalográfico.
15. Se ha resuelto la estructura cristalográfica del dominio catalítico Xyn10C-CD a 1,76 Å de resolución, con grupo espacial $P3_2$. Su estructura revela que Xyn10C-CD es capaz de unir xilanos altamente decorados, pudiendo alojar sustituciones en casi todos sus subsitios. Esta característica es común con lo observado en el sitio de unión de CBM22-2 y sugiere un modo de acción concertado entre ambos módulos, que pueden reconocer el mismo sitio de unión del sustrato, altamente heterogéneo.
16. En base a los resultados estructurales y funcionales, se ha formulado un posible papel del dominio N-terminal en Xyn10C. Así, CBM22-1 dirigiría la enzima a un primer sitio de reconocimiento del xilano, tras lo cual, CBM22-2 actuaría de manera más concertada con el dominio catalítico,

reconociendo sitios de unión específicos que van a ser degradados. El dominio C-terminal Xyn10C-CBD mantendría el contacto de la enzima con las microfibrillas de celulosa, próximas al xilano en la pared celular vegetal.

4.3. Artículo 3:



**Crystallization and preliminary X-ray
diffraction analysis of Xyn30D from
*Paenibacillus barcinonensis***

**María Ángela Sainz-Polo, Susana Valeria Valenzuela, F. Javier Pastor
and Julia Sanz-Aparicio**

Acta Crystallographica Section F (2014). **F70**, 963–966

4.3.1. Introducción a la cristalización de Xyn30D de *Paenibacillus barcinonensis*

La degradación enzimática de la pared celular vegetal es un proceso que cada día cobra mayor interés debido a sus implicaciones en el uso de fuentes de carbono renovables. Uno de los componentes mayoritarios de la pared celular vegetal es el xilano, siendo su despolimerización clave en la biodegradación de los materiales de origen vegetal. La degradación de estos materiales genera compuestos de gran utilidad y valor añadido con importantes aplicaciones en biotecnología y biomedicina. Sin embargo, la complejidad del xilano requiere de la acción de múltiples enzimas que actúan sinérgicamente para conseguir su completa degradación, existiendo enzimas capaces de romper tanto la cadena principal del polímero, como sus ramificaciones.

La enzima Xyn30D pertenece a un grupo de nuevas xilanasas adscritas a la familia GH30 que, a diferencia de las xilanasas clásicas, son específicas de determinadas sustituciones del polímero degradando únicamente GX. Junto a su módulo catalítico, Xyn30D contiene un dominio auxiliar no catalítico perteneciente a la familia CBM35. Presenta una masa de 62 kDa (incluyendo una cola de histidinas en el extremo C-terminal). Hasta la fecha, únicamente una enzima bimodular de la familia GH30 ha sido identificada y caracterizada, pero ninguna cristalizada.

La obtención de cristales de proteína es el paso previo clave para la resolución estructural por difracción de rayos X y posterior análisis de la estructura. Para ello, deben cumplirse ciertas exigencias de pureza y estabilidad de la muestra.

En el presente artículo se describe la purificación, cristalización y los datos preliminares cristalográficos de la proteína completa bimodular Xyn30D (GH30-CBM35). La resolución estructural por cristalografía de rayos X de Xyn30D contribuirá a esclarecer su mecanismo de actuación así como el papel de los dominios auxiliares no catalíticos en este tipo de enzimas modulares microbianas. Todo ello permitirá avanzar en el conocimiento de la sinergia que existe entre las distintas enzimas que forman la maquinaria molecular responsable de la degradación de la biomasa.

Los objetivos de este trabajo han sido los siguientes:

1. Definir un protocolo de purificación de Xyn30D que permita obtener muestra de calidad suficiente para poder llevar a cabo los posteriores experimentos de cristalización.
2. Establecer las condiciones óptimas de cristalización y difracción de rayos X de Xyn30D.
3. Interpretación de los datos estadísticos de difracción para una posterior resolución estructural satisfactoria de Xyn30D.

La autora de esta tesis ha realizado:

- a. Cristalización de Xyn30D.
- b. Toma de datos de difracción en el sincrotrón ESRF de Grenoble, en la línea ID23-2.
- c. Procesado de los datos y valoración de su calidad para la posterior resolución estructural.

4.3.2. Publicación



crystallization communications

Acta Crystallographica Section F
Structural Biology
Communications

ISSN 2053-230X

Crystallization and preliminary X-ray diffraction analysis of Xyn30D from *Paenibacillus barcinonensis*

**María Ángela Sainz-Polo,^a
Susana Valeria Valenzuela,^b
F. Javier Pastor^b and Julia
Sanz-Aparicio^{a*}**

^aDepartment of Crystallography and Structural Biology, Institute of Physical Chemistry 'Rocasolano', Consejo Superior de Investigaciones Científicas, Serrano 119, 28006 Madrid, Spain, and ^bDepartment of Microbiology, Faculty of Biology, Universitat de Barcelona, Avenida Diagonal 643, 08028 Barcelona, Spain

Correspondence e-mail: xjulia@iqfr.csic.es

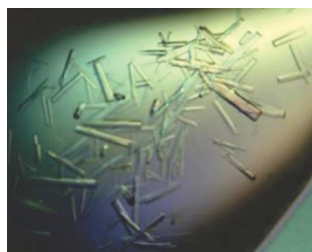
Received 4 April 2014
Accepted 23 May 2014

Xyn30D, a new member of a recently identified group of xylanases, has been purified and crystallized. Xyn30D is a bimodular enzyme composed of an N-terminal catalytic domain belonging to glycoside hydrolase family 30 (GH30) and a C-terminal family 35 carbohydrate-binding domain (CBM35) able to bind xylans and glucuronic acid. Xyn30D shares the characteristic endo mode of action described for GH30 xylanases, with the hydrolysis of the β -(1,4) bonds of xylan being directed by α -1,2-linked glucuronate moieties, which have to be placed at the –2 subsite of the xylanase active site. Crystals of the complete enzyme were obtained and a full data set to 2.3 Å resolution was collected using a synchrotron X-ray source. This represents the first bimodular enzyme with the domain architecture GH30-CBM35. This study will contribute to the understanding of the role that the different xylanases play in the depolymerization of glucuronoxylan.

1. Introduction

The enzymatic degradation of the plant cell wall is a process that is attracting more interest every day because of its implications for the use of renewable carbon sources. Xylan is one of the main components of the cell wall, which encompasses one third of the world's renewable carbon (de Vries & Visser, 2001). Consequently, its depolymerization plays a key role in the biodegradation of the plant cell wall. Xylanases are enzymes that catalyze the hydrolysis of the main chain of xylan and to date they have been successfully used in the textile, food, feed and paper industries among others. Nevertheless, the complexity of xylan requires the combined action of different xylanases and other enzymes able to break the heterogeneous side-chain substitutions present in highly branched xylans (Biely, 1985).

Commonly, xylanases are generally classified into the GH10 or GH11 families of glycosyl hydrolases, but in recent times new xylanases have been classified into families GH5, GH8, GH30 and GH43 (CAZy; <http://www.cazy.org/>, Lombard *et al.*, 2014). The GH30 family members share the common feature of only being active on glucuronoxylans, which is a unique trait of this xylanase family. Recently, the first two xylanases belonging to this family have been crystallized: glucuronoxylan xylanohydrolase from *Bacillus subtilis* (PDB entry 3kl0; St John *et al.*, 2011) and glucuronoxylanase from *Erwinia chrysanthemi* (PDB entry 2y24; Urbanikova *et al.*, 2011). Both enzymes show a catalytic domain with a $(\beta/\alpha)_8$ -barrel fold tightly connected to a side β -sheet structure. A new GH30 xylanase, *Paenibacillus barcinonensis* Xyn30D, has been identified and characterized (Valenzuela *et al.*, 2012). The enzyme is a bimodular enzyme comprised of a complete GH30 domain bound to an ancillary CBM35 carbohydrate-binding module (Montanier *et al.*, 2009; Correia *et al.*, 2010) and constitutes the first example of a modular xylanase within the GH30 family. Here, we report its purification, crystallization and preliminary X-ray crystallographic analysis.



© 2014 International Union of Crystallography
All rights reserved

Acta Cryst. (2014). F70, 963–966

doi:10.1107/S2053230X14012035 **963**

crystallization communications

The few existing reports on the catalytic properties of xylanases specific for glucuronoxylan hydrolysis make it difficult to identify common traits that can provide clues to understanding their contribution to the depolymerization of this polysaccharide. Further studies will be required to ascertain their role in the degradation of xylan in natural habitats. Elucidation of the Xyn30D structure will contribute to deciphering the biochemical function of the different xylanases present in a single microorganism.

2. Experimental procedures

2.1. Cloning and expression of Xyn30D

Construction of the Xyn30D expression vector for crystallization studies has been described previously (Valenzuela *et al.*, 2012). The vector expressed the full-length enzyme containing its own signal peptide and linked to a C-terminal His₆ tag. For the expression assay, *Escherichia coli* BL21 Star (DE3) was transformed with pET101Xyn30D and grown in ampicillin-supplemented LB medium (50 µg ml⁻¹) at 310 K until the absorbance at 600 nm reached 0.8; it was then induced with 0.5 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) for 18 h at 303 K. The cells were collected by centrifugation at 6000g for 15 min and resuspended in 50 mM Tris-HCl pH 7.0, 20 mM imidazole, 500 mM NaCl. Cells were disrupted with a French press. The solution was then centrifuged for 20 min at 20 000g and the resulting supernatant was placed on ice. Xyn30D was expressed with its own signal peptide and with its C-terminus fused to a 6×His tag to facilitate purification. The molecular mass of this construct is 62 124 Da.

2.2. Purification of Xyn30D

The recombinant Xyn30D was initially purified from cell extracts by immobilized metal-affinity chromatography (IMAC) using 1 ml HisTrap HP columns (GE Healthcare) and was eluted in 50 mM Tris-HCl pH 7.0, 500 mM NaCl with a 20–500 mM imidazole gradient on a fast protein liquid chromatography system (ÄKTA FPLC; GE Healthcare). The elution fractions were concentrated and dialyzed with 50 mM Tris-HCl pH 8.0 using Centricon centrifugal filter units

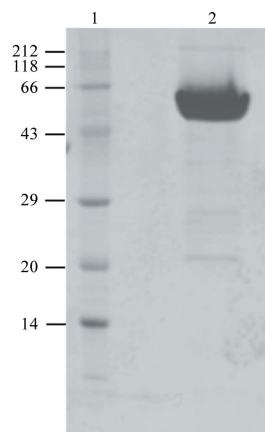


Figure 1
SDS-PAGE analysis of purified Xyn30D in 15% polyacrylamide gels. Lane 1, low-molecular-weight protein marker (labelled in kDa); lane 2, purified Xyn30D.

Table 1

Data-collection statistics for the Xyn30D crystal.

Values in parentheses are for the highest resolution shell.

| | |
|--|---|
| Wavelength (Å) | 0.873 |
| Source | ESRF |
| Beamline | ID23-2 |
| Space group | <i>P</i> ₃ 21 |
| Unit-cell parameters (Å) | <i>a</i> = <i>b</i> = 173.81, <i>c</i> = 183.74 |
| Resolution limits (Å) | 35.74–2.30 (2.42–2.30) |
| Unique reflections | 141869 (20540) |
| <i>R</i> _{merge} [†] | 0.149 (0.531) |
| Completeness (%) | 99.9 (100) |
| Mean multiplicity | 3.7 (3.7) |
| Mean <i>I</i> /σ(<i>I</i>) | 8.7 (2.8) |
| Wilson <i>B</i> factor (Å ²) | 17.04 |

[†] $R_{\text{merge}}(I) = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th measurement of a symmetry-related reflection with indices *hkl*.

of 30 kDa molecular-mass cutoff (Millipore) to a final volume of less than 10 ml. The concentrated enzyme was then loaded onto Mono Q 5/50 GL (GE Healthcare). Single injections of 1800 µl were made and the protein was eluted using 20 mM Tris-HCl pH 8.0 buffer with a gradient of 0–1 M NaCl. Purified Xyn30D was concentrated to at least 10 mg ml⁻¹ in 20 mM Tris-HCl pH 8, 150 mM NaCl for crystallization screening. The purity of the protein was determined by SDS-PAGE (Laemmli, 1970).

2.3. Enzyme assays

Xyn30D was assayed for enzyme activity using the method of Somogyi and Nelson (Valenzuela *et al.*, 2012) to detect the release of reducing sugar. Beechwood xylan (Sigma) at 1.5% (*w/v*) was included in the reactions as the substrate.

2.4. Crystallization

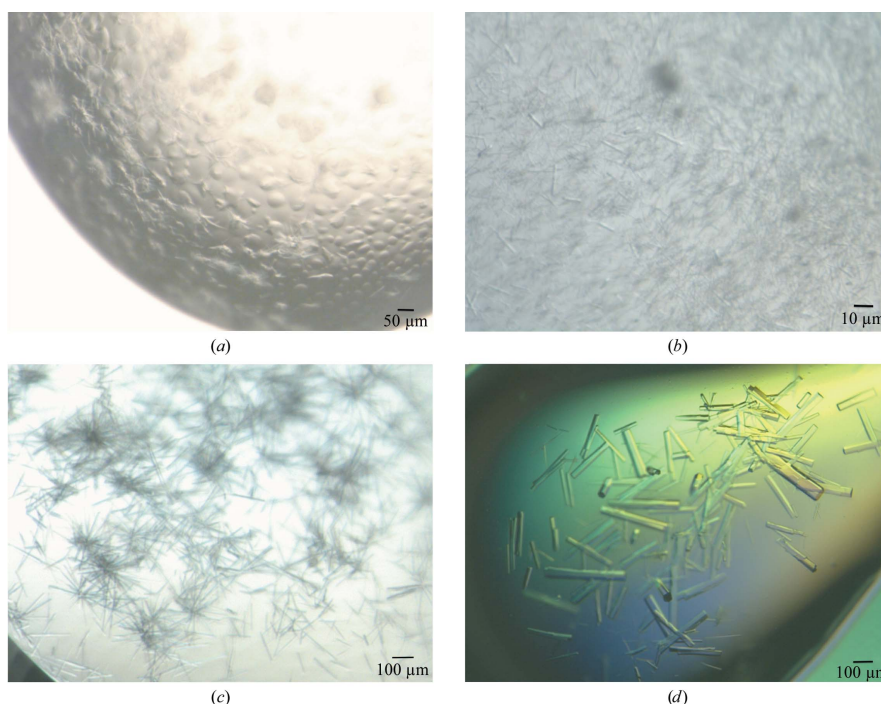
In order to determine a suitable protein concentration for crystallization trials, the Pre-Crystallization Test (PCT; Hampton Research) was used and 13 mg ml⁻¹ was found to be an appropriate value. Initial crystallization conditions of the purified protein samples were investigated by high-throughput techniques with a NanoDrop robot (Innovadyne Technologies Inc.) using the commercially available Index, SaltRx and Crystal Screen from Hampton Research and The PACT Suite and The JCSG+ Suite from Qiagen. The assays were carried out by the sitting-drop vapour-diffusion method at 291 K in Innovaplate SD-2 microplates (Innovadyne Technologies Inc.) by mixing 250 nl protein solution with 250 nl precipitant solution and equilibrating against 60 µl well solution.

A few PEG 3350-containing solutions gave small needle-like crystals. These conditions were optimized through further sitting-drop and hanging-drop experiments by mixing 2 µl protein solution with 1 µl precipitant solution and equilibrating against 500 µl well solution on VDX Plates (Hampton Research).

2.5. Data collection and processing

Crystals tested with synchrotron radiation were soaked in precipitant solution containing an additional 30% (*v/v*) glycerol (Garman & Mitchell, 1996) for a few seconds before being flash-cooled to 100 K. Diffraction data sets were collected on the ID23-2 beamline at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. The data sets were processed using *iMosflm* (Battye *et al.*, 2011) and *SCALA* (Evans, 2006) as distributed in the *CCP4* suite (Winn *et al.*, 2011).

crystallization communications

**Figure 2**

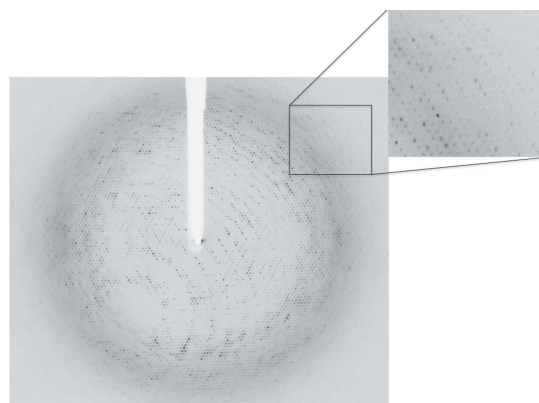
Xyn30D crystals obtained from (a) 15% (w/v) PEG 3350, 0.2 M magnesium chloride, 0.1 M Tris-HCl pH 8.5 (protein:precipitant ratio 1 μ l:1 μ l), (b) 14% (w/v) PEG 3350, 5% (v/v) glycerol, 0.1 M Tris-HCl pH 8.0 (protein:precipitant ratio 1 μ l:1 μ l), (c) 13% (w/v) PEG 3350, 5% (v/v) glycerol, 0.1 M Tris-HCl pH 8 (protein:precipitant ratio 2 μ l:1 μ l) and (d) 10% (w/v) PEG 3350, 5% (v/v) glycerol, 0.1 M Tris-HCl pH 8.0 and streak-seeding (protein:precipitant ratio 2 μ l:1 μ l).

3. Results and discussion

The purification of the Xyn30D sample through two chromatographic steps yielded an active protein with a high degree of purity (Fig. 1). A preliminary search for crystallization conditions with several commercial screens led to small needle-like crystals at 15% (w/v) PEG 3350, 0.2 M magnesium chloride, 0.1 M Tris-HCl pH 8.5 (Fig. 2a). Optimization was pursued manually by varying the protein/precipitant concentration and pH, and by trying different additives. The crystals in Fig. 2(b) were obtained when glycerol was included as an additive. Further improvements in the crystals (Fig. 2c) resulted from changing the drop ratio (to 2 μ l protein:1 μ l precipitant solution). Finally, streak-seeding (Bergfors, 2003) using the small needle crystals in Fig. 2(b) was crucial in obtaining high-quality crystals that allowed full data collection. Thus, the best rod-shaped crystals (Fig. 2d) were grown by streak-seeding into drops consisting of 2 μ l Xyn30D (13 mg ml⁻¹) in 20 mM Tris-HCl pH 8.0, 80 mM NaCl and 1 μ l of precipitant solution [10% (w/v) PEG 3350, 5% (v/v) glycerol, 0.1 M Tris-HCl pH 8.0]. The optimized crystals grew within 10 d at 291 K.

The final optimized crystals were tested using a synchrotron-radiation source to obtain high-resolution data. Several native data sets were collected at 100 K on the ID23-2 beamline at the ESRF, Grenoble, France. The crystals belong to space group $P3_221$, with unit-cell parameters $a = b = 173.81$, $c = 183.74$ Å, and diffracted to 2.3 Å resolution (Fig. 3, Table 1). As calculated from its sequence analysis, the molecular mass of the monomer is 62 124 Da. Assuming

a reasonable Matthews coefficient value within the range 4.87–2.44 Å³ Da⁻¹ (Matthews, 1968), corresponding to 50–70% solvent content, the presence of three to five molecules in the asymmetric unit should be expected. We investigated the local symmetry relating

**Figure 3**

X-ray diffraction pattern of Xyn30D obtained using a synchrotron source. The maximum observed resolution is 2.3 Å.

crystallization communications

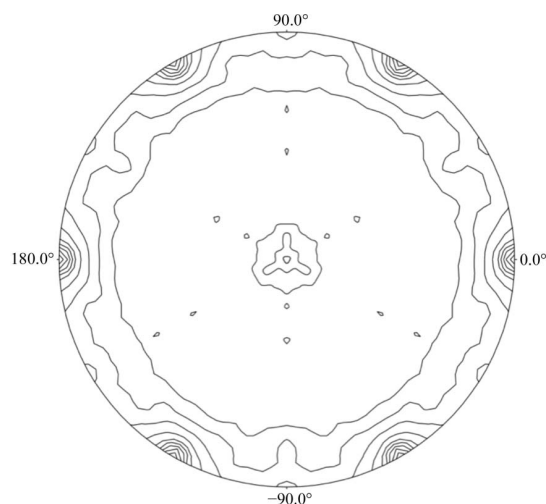


Figure 4
Plot of the self-rotation function of Xyn30D crystals in the $\kappa = 180^\circ$ section. The view is down the c axis and $\varphi = 0$ and $\varphi = 90^\circ$ correspond to the a and b^* axes, respectively.

the subunits in the asymmetric unit using *POLARREN* (Kabsch, 1976) from the *CCP4* package. Several self-rotation functions were computed in the resolution range 15–3.5 Å with Patterson vectors from 25 to 40 Å radius of integration. The stereographic projection of the $\kappa = 180^\circ$ section, shown in Fig. 4, revealed the presence of noncrystallographic twofold symmetry in the ac plane. Structure determination is in progress using two different models: firstly, the coordinates of glucuronoxylanase XynC from *B. subtilis* (St John *et al.*, 2009; PDB entry 3gtn), which shows 80% sequence identity to the GH30 portion of Xyn30D and, secondly, the C-terminal CBM35 from *Amycolaptosis orientalis* exo-chitosanase (Montanier *et al.*, 2009; PDB entry 2vzq), with 40% identity to the Xyn30D C-terminal domain. These models were used as templates in the molecular-replacement method with data to 3.5 Å resolution using *MOLREP*

(Winn *et al.*, 2011). A solution containing four molecules in the asymmetric unit has been found, which is consistent with twofold NCS symmetry. Preliminary structural refinement of this tetramer with *REFMAC* (Murshudov *et al.*, 2011) decreased the R factor to 0.41 ($R_{\text{free}} = 0.46$). Model building and further refinement are ongoing.

This work was supported by grants CTQ2010-20238-C03-02 and BIO2010-20508-C04-03 from the Dirección General de Investigación MINECO. This is a product of the Project 'Factoría Española de Cristalización' Ingenio/Consolider 2010. MASP is supported by a JAE-PreDoc fellowship from CSIC. We also thank the ESRF for beamtime and ID23-2 staff for providing assistance with data collection.

References

- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
 Bergfors, T. (2003). *J. Struct. Biol.* **142**, 66–76.
 Biely, P. (1985). *Trends Biotechnol.* **3**, 286–290.
 Correia, M. A. *et al.* (2010). *Biochemistry*, **49**, 6193–6205.
 Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
 Garman, E. F. & Mitchell, E. P. (1996). *J. Appl. Cryst.* **29**, 584–587.
 John, F. J. St, Godwin, D. K., Preston, J. F., Pozharski, E. & Hurlbert, J. C. (2009). *Acta Cryst.* **F65**, 499–503.
 John, F. J. St, Hurlbert, J. C., Rice, J. D., Preston, J. F. & Pozharski, E. (2011). *J. Mol. Biol.* **407**, 92–109.
 Kabsch, W. (1976). *Acta Cryst.* **A32**, 922–923.
 Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
 Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). *Nucleic Acids Res.* **42**, D490–D495.
 Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
 Montanier, C. *et al.* (2009). *Proc. Natl Acad. Sci. USA*, **106**, 3065–3070.
 Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
 Urbanikova, L., Vrsanska, M., Krogh, K. B., Hoff, T. & Biely, P. (2011). *FEBS J.* **278**, 2115–2116.
 Valenzuela, S. V., Diaz, P. & Pastor, F. I. (2012). *Appl. Environ. Microbiol.* **78**, 3923–3931.
 Vries, R. P. de & Visser, J. (2001). *Microbiol. Mol. Biol. Rev.* **65**, 497–522.
 Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.

4.3.3. Resultados y discusión

Xyn30D se expresó y purificó con una cola de histidinas en el extremo C-terminal. La purificación se llevó a cabo mediante dos pasos de cromatografía, afinidad (HisTrap) e intercambio iónico (Mono Q), tras lo cual, la muestra se dializó en tampón 20 mM Tris-HCl pH 8,0, 80 mM NaCl, para los experimentos de cristalización.

Un rastreo inicial de condiciones de cristalización se llevó a cabo con una extensa búsqueda de más 480 soluciones comerciales con la ayuda del robot *NanoDrop (Innovadine Technologies Inc.)* disponible en nuestro laboratorio. Una de dichas condiciones (15 % (m/v) PEG 3350, 0,2 M MgCl₂, 0,1 M Tris-HCl pH 8,5) generó unas pequeñas agujas. La estrategia de optimización de estos cristales preliminares fue variar la concentración de precipitante, la concentración de proteína, la proporción proteína/precipitante y el pH. Además, también se probaron un total de 96 aditivos. La adición de glicerol y el aumento de la proporción en la gota de proteína/precipitante (2:1) condujo a una mejora de esas agujas, obteniendo unas barras de buena calidad cristalina. Finalmente, y siguiendo una estrategia de “sembrado por arrastre”, se obtuvieron unas barras de tamaño adecuado para los experimentos de difracción. La concentración de proteína óptima fue de 13 mg ml⁻¹ y las condiciones finales de cristalización: 10 % (m/v) PEG 3350, 5 % (m/v) glicerol, 0,1 M Tris-HCl pH 8,0. Los mejores cristales salieron después de 10 días, a 291 K de temperatura.

Para crioproteger estos cristales se sumergieron en una solución con las condiciones de cristalización más un 30 % (m/v) de glicerol, tras lo cual se sumergieron en nitrógeno líquido a 100 K. Se usó radiación sincrotrón para obtener datos a alta resolución. Varios conjuntos de datos se recogieron a 100 K en la línea ID23-2 del ESRF de Grenoble, Francia. Los cristales pertenecen al grupo espacial *P3₂21*, con parámetros de celdilla $a = b = 173.81 \text{ \AA}$, $c = 183.74 \text{ \AA}$ y difractaron a una resolución de 2,3 Å. De acuerdo con el coeficiente de Matthews, dentro del intervalo 4,87–2,44 Å³ Da⁻¹, y correspondiente a un contenido de solvente de 50–70 %, se espera la presencia de 3 a 5 moléculas en la unidad asimétrica. Investigando la simetría local, es muy probable que existan 4 moléculas en la unidad asimétrica, siendo esto consistente con la simetría no cristalográfica.

4.4. Artículo 4:



Structural analysis of glucuronoxylan-specific Xyn30D and its attached CBM35 domain gives insights into the role of modularity in specificity

María Ángela Sainz-Polo, Susana Valeria Valenzuela, Beatriz González, F. I. Javier Pastor and Julia Sanz-Aparicio

The Journal of Biological Chemistry (2014) **289**, 45: 31088–31101

4.4.1. Introducción al análisis estructural de la glucuronoxilanasas Xyn30D y de su módulo auxiliar CBM35 como determinante de la especificidad

El hecho de que cada vez exista una mayor restricción en las leyes medioambientales ha llevado a que se busquen nuevas soluciones para hacer los procesos industriales más sostenibles. En este contexto se sitúa el incremento de la demanda para la utilización de enzimas en la degradación de biomasa, cuyos productos constituyentes tienen un gran valor añadido.

Las xilanasas son las enzimas que catalizan la hidrólisis del xilano, el cual es el componente mayoritario de la hemicelulosa. Este polímero está formado por una cadena de xilasas unidas por enlaces β -1,4, pudiendo estar sustituidos en las posiciones O2 y/o O3 de los anillos de xilosa. Por lo tanto, la composición del xilano es enormemente variable dependiendo del tipo de planta y tejido. Por su parte, el término glucuronoxilano se emplea con frecuencia para describir xilanos de maderas duras (con sustituyentes de ácido MeGlcA), mientras que GAX hace referencia a xilanos presentes en hierbas y cereales que presentan, además, gran cantidad de residuos de arabinosa. De este modo, la biodegradación del xilano requiere de la cooperación de múltiples enzimas coordinadas (GH) que son capaces de romper tanto la cadena principal como las ramificaciones. La mayoría de las xilanasas que se han caracterizado pertenecen a las familias GH10 y GH11, que en general tienen preferencia por sitios de baja ramificación. Sin embargo, cada vez se identifican más xilanasas con afinidad específica por un tipo concreto de xilano. Las glucuronoxilanasas, por ejemplo, son enzimas que requieren para su actividad de la existencia de sustituciones de MeGlcA, y se agrupan dentro de la familia GH30.

Muchas GH son enzimas modulares que presentan CBM anexos al módulo catalítico a través de segmentos conectores. Los CBM encontrados en xilanasas son capaces de reconocer no solo sustratos de las mismas, sino también otros como celulosa, en estrecha proximidad al xilano dentro de la pared celular de la planta. Estudios previos realizados con dominios de la familia CBM35 han sugerido que podrían dirigir las enzimas hacia regiones de la pared celular que estén siendo degradadas por agentes patógenos.

La glucuronoxilanasas Xyn30D pertenece al secretoma de *P. barcinonensis*, un microorganismo con un amplio conjunto de carbohidrolasas, incluyendo varias xilanasas. Del mismo modo que en los pocos ejemplos de xilanasas descritas dentro de la familia GH30, esta enzima tiene como requisito fundamental para la catálisis la presencia de ramificaciones de metil glucurónico y no es activa sobre AX. Una diferencia a destacar respecto a las GH30 ya caracterizadas estructuralmente (las glucuronoxilanasas de *Bacillus subtilis* y *Erwinia chrysanthemi*), es que estas dos enzimas son de dominio único mientras que Xyn30D es una enzima modular que posee un CBM35, raramente encontrado en xilanasas. Solamente otra xilanasas GH30 modular de *Clostridium*

papirosolvans ha sido secuenciada y la estructura de su módulo catalítico resuelta por cristalografía.

La proteína Xyn30D fue expresada, purificada y cristalizada previamente. Los objetivos del presente trabajo fueron continuar con el estudio de dicha enzima, y se recogen a continuación:

1. Resolución estructural de Xyn30D y de sus complejos.
2. Análisis estructural de sus dominios aislados.
3. Definir un protocolo de expresión y purificación del dominio auxiliar Xyn30D-CBM35.
4. Ensayos de unión con sustratos específicos mediante geles de afinidad de Xyn30D-CBM35.
5. Establecer sus condiciones óptimas de cristalización y difracción, con y sin ligandos.
6. Resolución estructural de Xyn30D-CBM35 y sus complejos. Interpretación de los datos estadísticos de los mismos.
7. Análisis de los datos estructurales y bioquímicos de Xyn30D, y de la contribución de Xyn30D-CBM35 en la despolimerización del xilano.

La autora de esta tesis ha realizado:

- a. Resolución estructural de Xyn30D.
- b. Cristalización de Xyn30D-CBM35 y de sus complejos Xyn30D-CBM35/GlcA y Xyn30D-CBM35/GlcA-Xyl.
- c. Toma de datos de difracción en varios sincrotrones europeos de las muestras cristalizadas en el apartado “b” y resolución estructural de las mismas.
- d. Interpretación de los resultados obtenidos y explicación del posible papel del dominio Xyn30D-CBM35.

4.4.2. Publicación

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 289, NO. 45, PP. 31088–31101, NOVEMBER 7, 2014
© 2014 BY THE AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY, INC. PUBLISHED IN THE U.S.A.

Structural Analysis of Glucuronoxylan-specific Xyn30D and Its Attached CBM35 Domain Gives Insights into the Role of Modularity in Specificity*

Received for publication, July 17, 2014, and in revised form, September 3, 2014. Published, JBC Papers in Press, September 8, 2014, DOI 10.1074/jbc.M114.597732

M. Angela Sainz-Polo⁺¹, Susana Valeria Valenzuela^{§2}, Beatriz González[‡], F. I. Javier Pastor[§], and Julia Sanz-Aparicio⁺³

From the [‡]Departamento de Cristalografía y Biología Estructural, Instituto de Química-Física Rocasolano, CSIC, 28006 Madrid, Spain and the [§]Departamento de Microbiología, Facultad de Biología, Universidad de Barcelona, 08028 Barcelona, Spain

Background: Xylanases are crucial in plant cell wall recycling.

Results: A glucuronoxylan-specific xylanase is attached to its binding module with moderate flexibility. This CBM35 displays novel structural features regulating specificity.

Conclusion: Depolymerization of highly substituted xyans and an oriented interaction with its target substrate are proposed.

Significance: Unraveling the mechanisms ruling modularity is essential to understanding the biomass deconstruction and to producing efficient biocatalysts.

Glucuronoxylanase Xyn30D is a modular enzyme containing a family 30 glycoside hydrolase catalytic domain and an attached carbohydrate binding module of the CBM35 family. We present here the three-dimensional structure of the full-length Xyn30D at 2.4 Å resolution. The catalytic domain folds into an (α/β)₈ barrel with an associated β -structure, whereas the attached CBM35 displays a jellyroll β -sandwich including two calcium ions. Although both domains fold in an independent manner, the linker region makes polar interactions with the catalytic domain, allowing a moderate flexibility. The ancillary Xyn30D-CBM35 domain has been expressed and crystallized, and its binding abilities have been investigated by soaking experiments. Only glucuronic acid-containing ligands produced complexes, and their structures have been solved. A calcium-dependent glucuronic acid binding site shows distinctive structural features as compared with other uronic acid-specific CBM35s, because the presence of two aromatic residues delineates a wider pocket. The nonconserved Glu¹²⁹ makes a bidentate link to calcium and defines region E, previously identified as specificity hot spot. The molecular surface of Xyn30D-CBM35 shows a unique stretch of negative charge distribution extending from its binding pocket that might indicate some oriented interaction with its target substrate. The binding ability of Xyn30D-CBM35 to different xyans was analyzed by affinity gel electrophoresis. Some binding was observed with rye glucuronoarabinoxylan in presence of calcium chelating EDTA, which would indicate that Xyn30D-CBM35 might establish interaction to other components of xylan, such as arabinose decorations of glucuronoara-

binoxylan. A role in depolymerization of highly substituted chemically complex xyans is proposed.

The utilization of biomass as a renewable source of biofuels, chemicals, and added value products has increasing economical and environmental relevance (1, 2). Biomass is composed mainly of plant cell walls, which contain a mixture of lignin and polysaccharides interlocked in a complex lignocellulose matrix. Among cell wall polysaccharides, xylan accounts for approximately one-third of the renewable organic carbon on earth (3, 4). It is composed of a backbone of β -1,4-xylose residues, which can be decorated at O2 with 4-O-methyl-D-glucuronic acid and at O2 or O3 with arabinofuranose residues, which can also be esterified to ferulic acid. Additionally, xylan can also be extensively acetylated. The composition of xylan is highly variable depending on the plant source and tissue, and the term glucuronoxylan is frequently used to describe hardwood xyans (highly substituted with 4-O-methyl-D-glucuronic acid residues), whereas (glucurono)arabinoxylan denominates xyans from grasses and cereals (with a large amount of arabinose residues) (4, 5). Biodegradation of xylan requires the coordinated activity of several enzymes, among which xylanases (1,4- β -D-xylan xylanohydrolase; EC 3.2.1.8) catalyze the cleavage of internal linkages on the xylose backbone. They play a central role in xylan depolymerization and are key enzymes for xylan bioconversion to fermentable sugars and oligomers of potential value as chemicals and prebiotics (6, 7). Biorefinery approaches to deconstruct the plant cell wall for the production of bioethanol and biomaterials have shown the relevance of improving cellulose accessibility to cellulases (8) and that removal of hemicelluloses can be more important than lignin removal (9). This evidences the important contribution of xylanases to the production of cellulosic ethanol from lignocelluloses with low lignin content, such as corn stover and other agricultural residues (10).

* This work was supported by Grants BIO2010-20508-C04-03 and CTQ2010-20238-C03-02 from the Spanish Ministry of Science and Innovation, Grant 2009-SGR-819-AGAUR from Generalitat de Catalunya, and by the BioStruct-X program through the European Commission.

The atomic coordinates and structure factors (codes 4QAW, 4QB1, 4QB2, and 4QB6) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

¹ Recipient of a predoctoral fellowship from the Spanish National Research Council (CSIC).

² Recipient of a doctorate grant from the Spanish Ministry of Foreign Affairs.

³ To whom correspondence should be addressed. Tel.: 34-91-561-9400; Fax: 34-91-564-2431; E-mail: xjulia@iqfr.csic.es.

The chemical complexity and heterogeneity of xylan can account for the multiplicity of xylanases produced by microorganisms (3, 11). They are grouped into glycoside hydrolase families based on amino acid sequence homologies and structural features within the CAZy database (12). Most characterized xylanases belong to families GH10 and GH11. They do not seem to be specialized for hydrolysis of a particular type of xylan, because they are able to degrade hardwood glucuronoxylans and arabinoxylans (11, 13). A small group of bacterial xylanases with specificity for glucuronoxylan has been identified and characterized. These enzymes, which require methylglucuronic decorations for activity, belong to subgroup H of the GH30 family and are considered as glucuronoxylanases (14–16). Opposite to substrate specificity shown by GH30 glucuronoxylanases, an arabinoxylan-specific xylanase has been reported (17). The enzyme belongs to family GH5 and seems the only reported example of xylanase with specificity for arabinoxylan. The occurrence of a variety of xylanases with differences in the substrate specificity and mode of action undoubtedly reflects the heterogeneity of plant xylan in natural habitats and contributes to its efficient degradation and utilization.

Glycoside hydrolases are frequently modular enzymes that contain catalytic modules joined by flexible linker sequences to carbohydrate binding modules (CBMs).⁴ They direct the appended catalytic modules to their target substrates and potentiate the activity against the complex substrates of cell wall. Similar to catalytic modules, CBMs are also classified into families based on amino acid sequence homologies (CAZy (18, 19)). Many xylanases are modular enzymes containing one or more CBMs. These noncatalytic modules found in xylanases belong to different families and recognize a diversity of ligands, including xylans, xylose oligomers, and carbohydrates that are not substrates of the xylanases, such as cellulose, in close proximity to xylan in the plant cell wall (18, 19). Analysis of family 35 CBMs suggested that they could direct the enzymes toward regions of the cell wall that are being actively degraded by pathogens (20). Additionally, the disruption of the substrate structure by CBMs of different families has been proposed in several reports (21, 22). In addition to getting the catalytic modules in close proximity to their substrates, CBMs probably have a more complex role in polysaccharides degradation, with additional functions to enable the deconstruction of cell wall and the catalytic depolymerization of carbohydrates.

Glucuronoxylanase Xyn30D belongs to the secretome of *Paenibacillus barcinonensis* (23), a powerful xylan degrading microorganism that shows a complex set of carbohydratases, including several GH10 and GH11 xylanases cloned and characterized (24–26). Like the few examples of characterized glucuronoxylanases of family GH30, the enzyme shows requirement of methylglucuronic substitutions for catalysis and is not active on arabinoxylans (27). However, unlike the structurally characterized GH30 glucuronoxylanases from *Bacillus subtilis* and *Erwinia*

Structure of Bimodular *P. barcinonensis* Xyn30D

chrysanthemi (28, 29), which are single domain enzymes, Xyn30D is a modular enzyme. The glucuronoxylanase contains a carbohydrate binding module of the CBM35 family, which is rarely found in xylanases. Only recently a modular GH30 xylanase from *Clostridium papirosolvans*, containing a CBM6, has been sequenced and its catalytic domain crystallized, although the biochemical characterization of the enzyme has not been reported (PDB code 4FMV). In our study, we have purified and crystallized glucuronoxylanase Xyn30D and analyzed the three-dimensional structure of the full-length enzyme. Its ancillary CBM35 domain has also been expressed and crystallized, and its binding abilities have been investigated by soaking experiments. Additionally, binding of purified CBM35 to soluble xylans has been analyzed by affinity gel electrophoresis. Xyn30D has a unique GH30-CBM35 modular assembly. The results here presented contribute to deciphering the biochemical function of GH30 xylanases and the contribution of CBM35 to the efficiency of xylan depolymerization. Further studies will be required to ascertain the role of glucuronoxylan-specific xylanases and the function of appended CBM in degradation of xylan in natural habitats.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification—Construction of Xyn30D, Xyn30D-GH30, and Xyn30D-CBM35 expression vectors were previously described (27). The protein samples were purified from *Escherichia coli* BL21Star (DE3) recombinant cultures containing plasmids pET101Xyn30D, pET101Xyn30D-GH30, and pET28aXyn30D-CBM35, respectively. Exponential phase cultures (A_{600} of 0.8) were induced with 0.5 mM isopropyl- β -D-thiogalactopyranoside for 18 h at 303 K. Cells were collected and disrupted by French press. Concentrated extracts were subjected to immobilized metal affinity chromatography using HisTrap HP columns of 1 ml (GE Healthcare), on a fast protein liquid chromatography system (AKTA FPLC; GE Healthcare). Washing with 50 mM Tris-HCl, pH 8, 50 mM imidazole, and 500 mM NaCl was performed before an elution in a single step with 50 mM Tris-HCl buffer of pH 8, 500 mM NaCl, and 300 mM imidazole. The immobilized metal affinity chromatography elution fractions were concentrated with Centricon centrifugal filter units of 3-kDa molecular mass cutoff (Millipore) and loaded for a second polishing step in a gel filtration column (Tricorn Superdex 200 10/300 GL; GE Healthcare). Single injections of 1,800 μ l were made, and the protein was eluted with 20 mM Tris-HCl, pH 8, 150 mM NaCl. The purity of the protein was verified by SDS-PAGE (30).

Crystallization and Data Collection—Crystals of Xyn30D were grown as described before (31). For data collection, native crystals were transferred to cryoprotectant solutions consisting of mother liquor plus 30% (v/v) glycerol before being cooled to 100 K in liquid nitrogen. Diffraction data were collected using synchrotron radiation at the European Synchrotron Radiation Facility (Grenoble, France) on ID23-2 Beamline. Diffraction images were processed with iMOSFLM (32) and merged using the CCP4 package (33). Crystals of Xyn30D were indexed in the $P3_22_1$ space group with four molecules in the asymmetric unit and 60% solvent content within the unit cell.

⁴ The abbreviations used are: CBM, carbohydrate binding module; a.u., asymmetric unit; GlcA, glucuronic acid; GH, glycosylhydrolase; MeGA, methylglucuronic acid; MeGX, methylglucuronic acid-substituted xylooligosaccharide; NCS, noncrystallographic symmetry; RMS, root mean square; PDB, Protein Data Bank.

Structure of Bimodular *P. barcinonensis* Xyn30D

Crystals from Xyn30D-CBM35 isolated module were obtained by mixing 1 μ l of 22 mg/ml protein solution (20 mM Tris, pH 8.0, and 150 mM NaCl) with 1 μ l of a solution containing 20% (w/v) PEG 6000, 0.2 M Ca_2Cl , and either 0.1 M MES, pH 6.0, or 0.1 M HEPES, pH 7.0, and equilibrating by vapor diffusion at room temperature. Complexes of Xyn30D-CBM35 were obtained by the soaking technique using glucuronic acid (GlcA; Sigma, or aldouronic acid mixture purchased from Megazyme (containing aldotriouronic, aldotetraouronic, and aldopentaouronic acids, 2:2:1). The crystals were soaked for 30 min in solutions made of mother liquor plus 30 mM GlcA or a 1/100 dilution of the commercial aldouronic mixture. For data collection, all crystals were transferred to cryoprotectant solutions containing mother liquor supplemented with 20% (v/v) glycerol before being cooled to 100 K in liquid nitrogen. Crystals showing diverse habits belonged to $\text{P}2_12_12_1$ space group with one molecule in the asymmetric unit and 55% solvent content within the unit cell. Diffraction data for Xyn30D-CBM35 were collected in-house with a MAR345dtb (MarResearch) detector equipped with a rotating anode generator (MicroStar, Bruker) and Helios mirrors. X-ray data from the complexes were collected using different synchrotron sources at Desy (Germany) and Diamond (UK), on the beamlines given in Table 1. Diffraction images were processed with iMOSFLM (32) and XDS (34) and merged using the CCP4 package (33). A summary of data collection and data reduction statistics is shown in Table 1.

Structure Solution and Refinement—The structure of Xyn30D was solved by molecular replacement using the MOLREP program (36). The structures of glucuronoxylanase XynC from *B. subtilis* (PDB code 3GTN) and CBM35 from *Amycolaptosis orientalis* exo-chitosanase (PDB code 2VZQ) were used to prepare the search model using the program Chainsaw (37) and a protein sequence alignment of Xyn30D onto both templates. A partial solution containing four molecules of the Xyn30D catalytic domain in the asymmetric unit (a.u.) was found using reflections up to 3.0 Å resolution range and a Patterson radius of 40 Å. Fixing these partial model, a solution containing one additional Xyn30D-CBM35 molecule was found using a Patterson radius of 15 Å. Then a complete Xyn30D molecule was generated and used as template with a Patterson radius of 52 Å, finally leading to a solution including all the four Xyn30D molecules, which after rigid body fitting led to an *R* factor of 49%. Crystallographic refinement was performed using the program REFMAC (38) within the CCP4 suite with flat bulk solvent correction, maximum likelihood target features, and local non-crystallographic symmetry (NCS). Free *R* factor was calculated using a subset of 5% randomly selected structure-factor amplitudes that were excluded from automated refinement. Model building using the program COOT (39) was combined with several rounds of refinement leading to a model showing a continuous density for the whole polypeptide chain, excluding the linker region. However, the refinement was stuck at an *R* factor of 38% ($R_{\text{free}} = 39$), and lowering space group symmetry to $\text{P}3_2$, with eight independent molecules in the a.u., was necessary to allow further progress to an acceptable agreement (*R* factors of 25/29%). Furthermore, two NCS groups composed of the catalytic GH30 and the CBM35 moieties, respectively, were defined, and the refinement was accomplished applying

medium NCS restraints. After iterative refinement and rebuilding of the linker regions, the final $2F_o - F_c$ map showed continuous density for the whole protein in chains A, B, and H, whereas some regions of the linker were not visible in chains C (amino acids 399–400), D/G (amino acids 398–400), and E/F (amino acids 397–401). At the later stages, water molecules were included in the model, which, combined with more rounds of restrained refinement, led to a final *R* factor of 17.9 ($R_{\text{free}} = 20.4$) for all data set up to 2.4 Å resolution. Refinement parameters are reported in Table 1.

The structure of Xyn30D-CBM35 was solved by molecular replacement using MOLREP (36) and the coordinates of the 403–530 portion of Xyn30D as the search model. Crystallographic refinement was performed using REFMAC (38) combined with model building with COOT (39) and addition of water molecules, which led to a final *R* factor of 20.9 ($R_{\text{free}} = 26.9$). The structure of Xyn30D-CBM35 complexed with GlcA and aldouronic acid was solved by difference Fourier synthesis using these refined coordinates. The ligands were manually built into the electron density maps and were refined similarly, to reach the *R* factors listed in Table 1.

Stereochemistry of the models was checked with PROCHECK (40) and MolProbity (41). The figures were generated with PyMOL (42). RMS deviation analysis was done using the program SUPERPOSE within the CCP4 package (33). Coordinates for all the structures have been deposited in the Protein Data Bank under accession numbers 4QAW, 4QB1, 4QB2, and 4QB6.

Polysaccharide Chain Model Building—A PDB model of a portion of the glucuronoarabinoxylan chain, with a typical substitution pattern (5) was constructed using the online biomolecule building program GLYCAM (43) and exported in its lowest energy state. The xylan chain was modeled into the Xyn30D-CBM35 binding cleft by superimposition of its GlcA-xylose moiety onto the experimental position observed in the aldouronic acid soaked crystals. Then only small rotations of the glycoside bond were introduced manually in the two last xylose units at the reducing term of the xylooligosaccharide chain, to fit the ligand model to the Xyn30D-CBM35 binding cleft surface. The xylooligosaccharide conformation obtained in this way keeps the reported typical xylan 3-fold helix pattern. Furthermore, the five central xylose units are in a conformation most similar to that experimentally observed in xylopentaose bound to *Pseudomonas fluorescens* xylanase A (PDB code 1EZN) (44).

Affinity Gel Electrophoresis—AGE was performed by following the method by Correia *et al.* (45) with some modifications. Continuous native polyacrylamide gels containing 6% acrylamide in 25 mM Tris, 200 mM glycine buffer (pH 8.3) were used. Soluble xylan (5–7 mg/ml) from beechwood (Roth), hardwoods (Sigma), or rye (Megazyme) was included in gels before polymerization. Gels with and without xylan were polymerized at the same time and were run in the same gel tank. Approximately 6 μ g of target protein was loaded in each well at room temperature, and gels were run at 30 mA/gel for 1 h. 15 mM EDTA, pH 8, was added to the samples when indicated. BSA was used as a negative noninteracting control.

TABLE 1
Crystallographic statistics

Values in parentheses are for the high resolution shell.

| Crystal data | Xyn30D | Xyn30D-CBM35 | Xyn30D-CBM35/GlcA | Xyn30D-CBM35/GlcA-Xyl |
|--|------------------------|------------------------|-------------------------|------------------------|
| Space group | $P3_2$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ |
| Unit cell parameters | | | | |
| <i>a</i> (Å) | 174.0 | 39.4 | 39.9 | 39.8 |
| <i>b</i> (Å) | 174.0 | 47.4 | 47.5 | 47.3 |
| <i>c</i> (Å) | 183.9 | 103.0 | 102.7 | 103.0 |
| Data collection | | | | |
| Beamline | ID23-2 (ESRF) | Rotating anode | P13 (PETRA III/DESY) | I03 (Diamond) |
| Temperature (K) | 100 | 100 | 100 | 100 |
| Wavelength (Å) | 0.8726 | 1.5418 | 0.9786 | 0.9762 |
| Resolution (Å) | 40.76-2.40 (2.53-2.40) | 51.51-2.18 (2.30-2.18) | 102.74-1.50 (1.58-1.50) | 34.86-1.35 (1.42-1.35) |
| Data processing | | | | |
| Total reflections | 1,435,305 (209,187) | 123,458 (14,761) | 414,192 (60,119) | 502,825 (73,033) |
| Unique reflections | 125,591 (18,229) | 10,371 (1,270) | 32,026 (4,619) | 43,659 (6,256) |
| Multiplicity | 11.4 (11.5) | 11.9 (11.6) | 12.9 (13.0) | 11.5 (11.7) |
| Completeness (%) | 100.0 (100.0) | 97.5 (83.8) | 99.6 (100) | 100 (100) |
| <i>I</i> /σ(<i>I</i>) | 4.8 (1.4) | 8.8 (4.2) | 4.4 (1.5) | 4.6 (1.5) |
| Mean <i>I</i> /σ(<i>I</i>) | 14.7 (4.6) | 25.7 (12.9) | 16.0 (5.1) | 17.3 (6.6) |
| <i>R</i> _{merge} (%) ^a | 14.3 (57.7) | 6.8 (16.9) | 9.7 (51.0) | 8.4 (42.7) |
| <i>R</i> _{pim} (%) ^b | 4.4 (17.8) | 2.0 (5.1) | 2.8 (14.6) | 2.6 (13.0) |
| Molecules per ASU | 8 | 1 | 1 | 1 |
| Refinement | | | | |
| <i>R</i> _{work} / <i>R</i> _{free} (%) ^c | 17.9/20.4 | 20.9/26.9 | 15.7/16.7 | 16.4/17.5 |
| No. of atoms/average B (Å ²) | | | | |
| Protein | 33,155/23.4 | 976/17.2 | 1,060/15.4 | 1,149/11.3 |
| Carbohydrate | 0 | 6/17.8 | 26/15.3 | 21/24.4 |
| Calcium ions | 16/36.1 | 2/15.4 | 2/11.9 | 2/8.3 |
| Water molecules | 95/14.9 | 56/20.5 | 121/24.4 | 116/22.1 |
| Ramachandran plot (%) | | | | |
| Favored | 96 | 98 | 98 | 99 |
| Outliers | 0.5 | 0 | 0 | 0 |
| RMS deviations | | | | |
| Bonds (Å) | 0.0073 | 0.0091 | 0.0149 | 0.0138 |
| Angles (°) | 1.2439 | 1.2904 | 1.4939 | 1.5773 |
| PDB accession codes | 4QAW | 4QB1 | 4QB2 | 4QB6 |

^a $R_{\text{merge}} = \sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl})$, where $I_i(\text{hkl})$ is the *i*th measurement of reflection *hkl* and $\langle I(\text{hkl}) \rangle$ is the weighted mean of all measurements.^b $R_{\text{pim}} = \sum_{\text{hkl}} [1/(N-1)] \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl})$, where *N* is the redundancy for the *hkl* reflection.^c $R_{\text{work}}/R_{\text{free}} = \sum_{\text{hkl}} |F_o - F_c| / \sum_{\text{hkl}} |F_o|$, where F_c is the calculated and F_o is the observed structure factor amplitude of reflection *hkl* for the working/free (5%) set, respectively.

RESULTS

Xyn30D Is a Bimodular Enzyme—We have purified and crystallized the glucuronoxylan-specific xylanase from *P. barcinonensis* (Xyn30D), as previously reported (31). We present here the three-dimensional structure of the full-length bimodular enzyme solved by molecular replacement at 2.4 Å resolution. Experimental and structure determination details are given under “Experimental Procedures” and in Table 1. An initial solution was obtained in the $P3_2$ space group containing four independent molecules in the asymmetric unit. Preliminary refinement with 4-fold NCS restraints allowed complete model building of the polypeptide chain, but full refinement was unfeasible until the space group was loosened to $P3_2$, which allowed the final parameters to converge to proper values. Therefore, the final model contains eight molecules in the a.u. presenting strong binary NCS along the cell axes. This feature is due to small differences in the orientation between the catalytic and its appended CBM domain found among the different molecules in the crystal, as it will be shown below. Each chain (A-B-C-D-E-F-G-H) contains the protein after cleavage of the signal peptide and consists of 537 residues with a molecular mass of 65 kDa as calculated from its primary structure. The model contains seven residues of the C-terminal polyhistidine tag (Ala⁵³¹–Ser⁵³⁷). An almost contiguous electron density was observed in three molecules of the a.u., only a small fragment of the linker region show-

ing poor or chopped density in the others, which precluded to fit a segment varying from one to five residues.

Fig. 1 illustrates the secondary structure and the molecular shape of Xyn30D. The enzyme is ~95 Å long by 55 Å wide by 40 Å thick. The GH30 catalytic domain has a CBM35 domain appended to its C terminus folded in an independent manner and is located almost perpendicularly and at the opposite part of the active center (Fig. 1b). Following the general pattern found in GH30 enzymes, Xyn30D catalytic domain folds into two subdomains. The major domain is a $(\alpha/\beta)_8$ barrel that starts at residue 11 and extends to residue 298. It has eight parallel β -strands that form the central barrel connected by eight external α -helices, the loops at the C-terminal end of the β -strands, L1–L8, contouring the catalytic pocket that is located in the center of the barrel. A tightly associated β -structure is fused to this barrel through a hydrophobic patch at α -helices 7 and 8. This second subdomain presents a nine-stranded β -sandwich, with immunoglobulin-like fold, composed of residues 1–10, at the N terminus, and residues 299–389, which is connected to the barrel through two segments with conserved sequence. This side β -structure is unique to the GH30 family and constitutes the main distinctive feature, as compared with the related GH5 enzymes (46).

The appended CBM35 noncatalytic domain (residues 402–530) displays a jellyroll type β -sandwich fold of two antiparallel sheets,

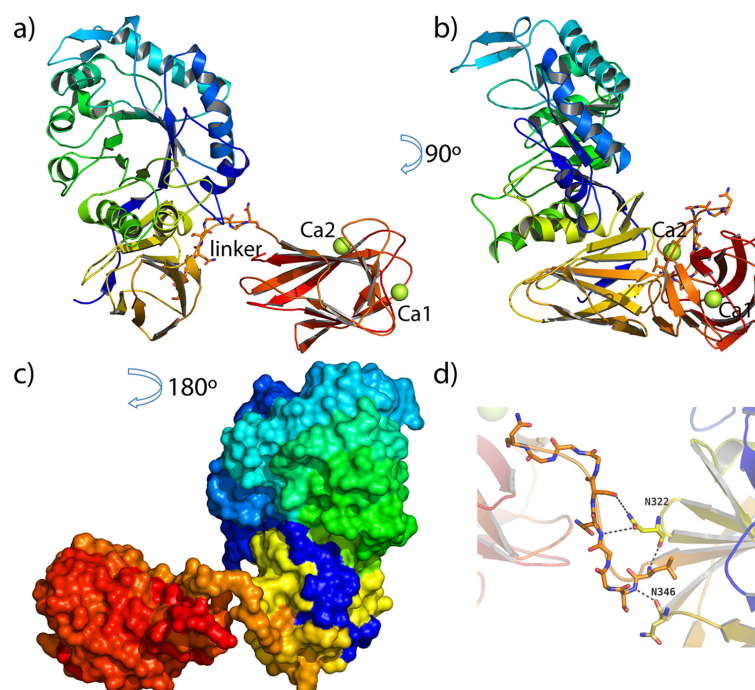
Structure of Bimodular *P. barcinonensis* Xyn30D

FIGURE 1. **The structure of Xyn30D.** *a*, view of Xyn30D showing its molecular architecture colored in rainbow sequence code, composed of a catalytic GH30 domain and an appended CBM35 domain. The catalytic domain folds into a $(\alpha/\beta)_8$ barrel and a tightly associated nine-stranded β -sandwich, connected to the barrel through two linkers with conserved sequence. The appended CBM35 domain displays a jellyroll type β -sandwich fold with two antiparallel sheets. Two calcium ions are represented as green spheres, whereas the linker segments are shown as sticks. *b* and *c*, the view is rotated 90° (*b*) and 180° (*c*), showing the molecular surface. *d*, zoom of the linker showing a net of hydrogen bonds that fixes this region to the β -structure of the GH30 domain.

formed by four and five antiparallel β -strands, respectively. Two calcium ions are observed: one is common to many lectins and CBM families having a structural role (Ca2), whereas the other (Ca1) is particular to CBM35 family and participates in substrate binding. An extended 12-residue segment, which is almost entirely visible in some molecules within the a.u., links this CBM35 to the catalytic GH30 domain. Although this region presents a nonconserved sequence typical of high mobility regions (LSGGNSGGGNVN), the first part of the linker (Leu³⁹⁰–Ser³⁹⁵) is fixed to the side β -structure by a net of polar interactions. As it is shown in Fig. 1*d*, Asn³²² and Asn³⁴⁶, from the second and fourth loop of the β -structure, respectively, are making several hydrogen bonds to the linker that fix this segment in a rather stable conformation conserved in all the eight independent molecules within the a.u. This feature possibly reduces motion to a short segment of the linker (Gly³⁹⁶–Asn⁴⁰¹), precluding the possibility of a large conformational repertoire. Consequently, only small differences in the orientation between both domains are observed in the crystal, going from 2° (molecules A *versus* C) to 5° (molecules A *versus* D) within the a.u. However, it must be taken into account that the linker segment Gly³⁹²–Gly³⁹³ is also involved in intermolecular packing contacts as explained below, and therefore, the linker might be more loosen in absence of these crystal restrictions. Nevertheless, and apart from this slight flexibility between

domains, no other significant differences among the eight independent molecules are observed in the crystal structure. Thus, superimposition of their catalytic domains gives a RMS deviation of 0.2–0.4 on 390 matched C α atoms, whereas superimposition of the CBM35 gives 0.2–0.3 on 127 matched C α atoms.

Despite many attempts, no structure of full-length Xyn30D in complex with substrate or reaction products has been obtained. Co-crystallization experiments did not yield crystals, whereas ligand-soaked Xyn30D crystals resulted in very poor diffraction. An inspection of the crystals shows that the linker region Gly³⁹²–Gly³⁹³ and also the side chain of Lys⁵³² at the beginning of the C-terminal polyhistidine flag are positioned in the substrate binding cleft of a contiguous molecule (Fig. 2). Consequently, diffusion of the ligands into the active site probably disrupts this interaction that must be important for stabilizing the intermolecular interface. In an attempt to overcome this problem, new constructions expressing the GH30 and the CBM35-independent domains were made, but only the CBM35 moiety provided suitable crystals for ligand soaking experiments.

The Catalytic Domain Resembles Known GH30 Xylanases— The three-dimensional structure has been reported for three other members of subgroup H within GH30 family, two glucuronoxylanases from *B. subtilis*, BsXynC (28), *E. chrysanthemi* EcXynA (47), and a glucuronarabinoxylanase from *C. papyrosolvens* (CpC71; PDB code 4FMV). Similarly to Xyn30D, CpC71

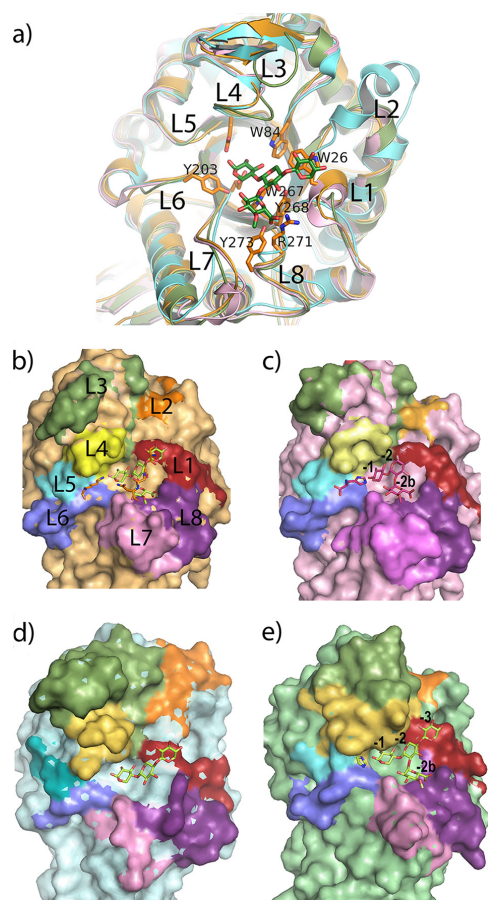


FIGURE 2. The structurally known GH30 xylanases. *a*, superimposition of Xyn30D (orange) to the other known GH30 enzymes belonging to subgroup H: glucuronoxylanases BsXynC (pink) and EcXynA (forest green), and glucuronoarabinoxylanase CpC71 (blue). MeGX3 ligand bound to the EcXynA active sites are shown as green sticks and used to extrapolate the putative ligand position in *b* and *d*. Some aromatic residues involved in substrate binding at loops L1, L3, L6, and L8 are indicated. The catalytic residues are also shown. *b*, molecular surface of Xyn30D highlighting L1–L8. The segment Ser³⁹¹–Asn³⁹⁴ from a symmetry-related molecule (orange) are located at the active site cleft, in a position similar to MeGX3 found in EcXynA (green), whereas Lys⁵³² from the same symmetry-related molecule is located at a putative site +1. *c*, molecular surface of BsXynC with MeGX2 and His found in the crystals at the active site. *d*, molecular surface of CpC71 with a Xyl3 putative chain at inferred position. *e*, molecular surface of EcXynA with the experimental position of the ligand.

has an attached noncatalytic CBM6 domain, but the reported crystal structure contains only the GH30 moiety.

Xyn30D shares the highest identity (80%) with BsXynC, whereas 57 and 40% identity are found with respect to CpC71 and EcXynA, respectively. However, the sequence conservation within the group is not equally distributed along the polypeptide chain, being significantly concentrated at the catalytic domain when compared with BsXynC (86%). Interestingly, CpC71 presents the highest identity with Xyn30D at the side

Structure of Bimodular *P. barcinonensis* Xyn30D

β -structure subdomain (73%). This trend reflects a differential role of this subdomain among the different enzymes that, in turn, must be related to the domain composition presented by each enzyme within the GH30 family, as will be explained below.

Structural superimposition of Xyn30D with the structurally known GH30 enzymes, shown in Fig. 2*a*, provides some insight into the specificity displayed by Xyn30D. The C α atoms of BsXynC, CpC71, and EcXynA structures superimposed onto the Xyn30D catalytic domain with RMS deviations of 0.6, 1.3, and 1.2, respectively (based on the overlap of 337, 319, and 317 matched C α atoms). As can be seen in the figure, larger differences are observed at loops surrounding the axis of the barrel, especially L1, L2, L3, and L4, which must account for the different substrate specificities. Fig. 2 (*b–e*) displays the molecular surface of the different enzymes and the involvement of each loop in shaping their active site clefts. Consistent with their endo mode β -1,4-xylanase activity, all enzymes present an extended crevice able to allocate the polymeric chain of xylan. Furthermore, GH30 family differs to classical GH10 and GH11 xylanases in that they are able to degrade heavily substituted xyans, this feature being reflected in quite open, exposed to solvent, binding clefts. The crystal structure of ligand-bound BsXynC (Fig. 2*c*) and EcXynA (Fig. 2*e*) have been also reported (28, 29). These two complex structures have shown how the MeGXn chain is attached to the active site cavity at the aglycon moiety of the substrate and allowed a detailed description of the different binding subsites. In particular, loops L1, L3, L6, and L8, present well conserved and exposed aromatic residues (Trp²⁶, Trp⁸⁴, Tyr²⁰³, Trp²⁶⁷, and Tyr²⁶⁸ in Xyn30D numbering), which fit the xylan chain (Fig. 2*a*). Most key determinants responsible for substrate binding at subsites –3 to –1 are also conserved in the known GH30 xylanases belonging to subfamily H, as is shown in the structural alignment given in Fig. 3. The main difference is the lack in CpC71 of the residues involved in the methylglucuronic acid (MeGA) recognition described previously in glucuronoxylanases, especially Arg²⁷¹ and Tyr²⁷³ from L8, both interacting with the C6 carboxylate, which are substituted by Trp²⁶⁹ and Asn²⁷¹ in CpC71. This ionic interaction with the uronic acid has been described as the main specificity determinant distinctive of the mode of action of bacterial GH30 xylanases (28, 29). CpC71 has been described as a glucuronoarabinoxylanase, but little has been reported on its particular functionality, which precludes the understanding of the structural basis of its dual substrate specificity. Apart from this feature, the CpC71 binding cleft topology is more similar to Xyn30D and BsXynC than that to that found in EcXynA, which is mainly due to the different conformation of loops L3 and L4. This is consistent to the two clusters described by phylogenetic analysis of the GH30 subgroup H enzymes corresponding to Gram-positive and Gram-negative bacteria (28).

Although soaking experiments of BsXynC and EcXynA crystals did not result in any ligand included at positively numbered subsites of the substrate, some unexpected molecules have been trapped in the crystals. Thus, a histidine, from the polyhistidine tag, and imidazol, from the buffer, occupy a position in the active site cleft of BsXynC and EcXynA that is similar to that of Lys⁵³² from the C terminus of a symmetry-related molecule found in Xyn30D crystals (Fig. 2). All these molecules are stack-

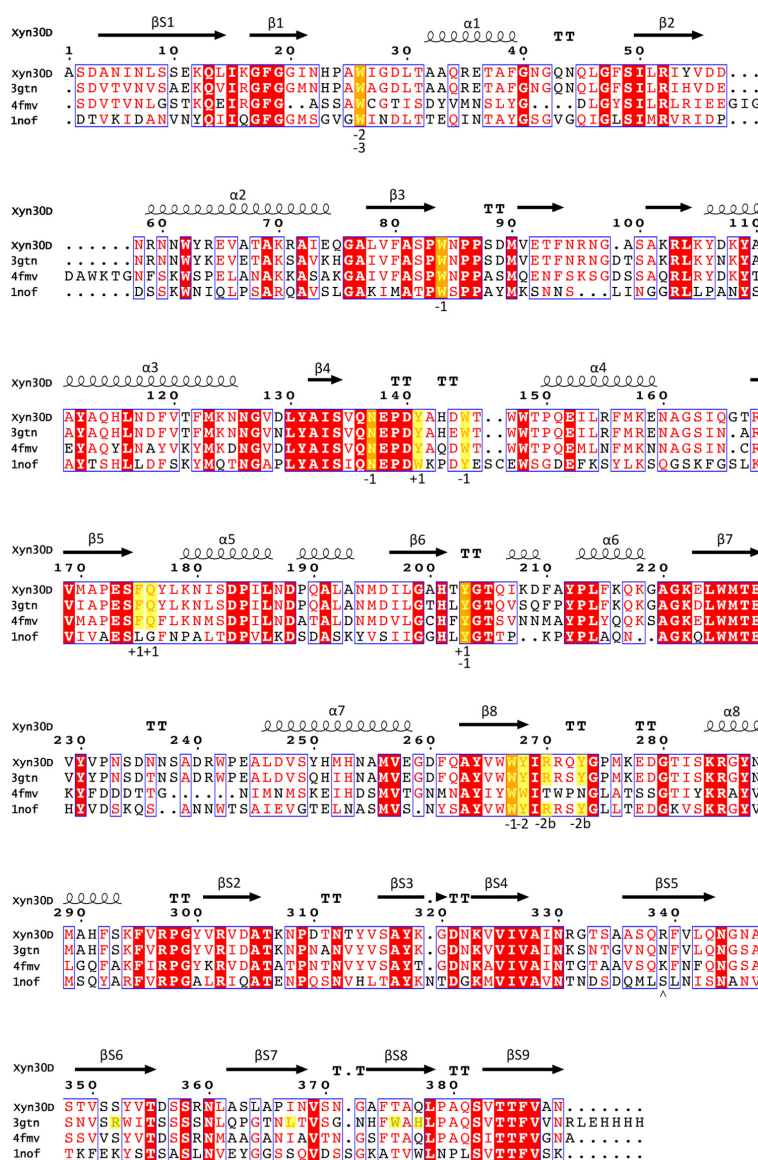
Structure of Bimodular *P. barcinonensis* Xyn30D

FIGURE 3. Structural alignment of GH30 subgroup H members (48). 3gtN, *B. subtilis* glucuronoxylanase; 4fmv, *C. papyrosolvens* glucuronarabinosylase; 1nof, *E. chrysanthemi* glucuronoxylanase. Relevant residues are highlighted in yellow.

ing to the conserved Tyr²⁰³, located at loop L6. There is a second aromatic residue Tyr/Trp at loop L4 (Fig. 3) that might be making additional stacking interaction to a putative xylopyranose ring occupying this subsite +1. The different chemical nature of the moieties bound in the different crystals suggests that binding to this subsite might be mostly controlled by hydrophobic stacking interaction, which possibly tolerates allocating different decorations of xylose at this position.

From all the structural determinants responsible for substrate binding found in subgroup H, only Tyr²⁰³ and Trp²⁶⁷ from L6 and L8 are conserved among GH30 enzymes. These two residues are making the hydrophobic floor that shapes subsites -1 and +1, within the active site cleft. This is not surprising taking into account that GH30 is a broad specificity and structurally diverse family, sharing less than 30% identity between its eight different subgroups. Moreover, GH30 sub-

families can be grouped within two main subgroups showing significant divergences as, for example, topological differences at its fused β -structure (46). Interestingly, a DALI (48) structural alignment identifies closest relatives to Xyn30D that include enzymes from families GH39, GH44, GH51, and GH79. These Clan-A GH families show identity levels to Xyn30D between 14 and 16%, analogous to the value of 19% when compared with human glucosylceramidase (PDB code 2V3F), which belongs to GH30 subgroup A. All four related GH families present the same $\alpha/\beta+\beta$ fold and contain enzymes that hydrolyze polysaccharides of lignocellulosic biomass.

The Substrate Binding Ability of the Fused β -Domain Is Not a Common Trend across GH30 Enzymes—Although the dual domain fold has been suggested to be a requirement for evolved evolution within GH30 (46), a role in xylan binding has been reported for the β -side structure in BsXynC (28). This observation came from soaking experiments with MeGX2 that showed a binding site for the ligand and a MeGA-specific recognition motif. Thus, the reducing-terminal xylose was stacking to Trp³⁷⁶ (BsXynC numbering), whereas Arg³⁵³ interacted to the MeGA moiety attached to the adjacent xylose ring (Fig. 4). Furthermore, Leu³⁶⁸ was making hydrophobic interaction with the methyl group, and this was proposed to discriminate MeGA from its unmethylated analogue. Although this specific binding motif is not conserved among subgroup H of GH30, putative MeGA binding has been also proposed for EcXynA (29). Thus, close to the site described in BsXynC, the EcXynA β -domain shows two candidates in the cleft that may keep the same function, *i.e.* an aromatic (Trp⁴⁰¹) and a basic residue (Lys³⁷⁵) able to stabilize a xylose ring and the carboxylate of its attached MeGA. These residues are all located in strands β S6, β S7, and β S8, which together with the precedent β S5 are the less conserved region of fused β -domain (Fig. 3).

An inspection to this subdomain in the Xyn30D structure shows two arginines, Arg³³⁹ and Arg³⁵⁹, that might be envisaged as potential candidates to target the MeGA carboxylate (Fig. 4a). Moreover, and close to Arg³⁵⁹, Leu³⁶⁴ could be properly orientated to make a hydrophobic environment to a putative adjacent methyl group. However, none aromatic residue is found on the surface that could target any attached xylose moiety. Consequently, the substrate binding ability observed in BsXynC and stated for EcXynA might not be a common trend of family GH30. In relation to this, it is remarkable that closest homologues to the Xyn30D β -fused subdomain (above 70% identity) present an appended separated CBM domain. Consequently, the specific role of the β -side structure seems to be depending on the domain composition of each GH30 enzyme.

The Xyn30D-CBM35 Domain Binds Glucuronic Acid—Xyn30D harbors a family 35 CBM at its C-terminal end, which folds into a jellyroll type β -sandwich typical of other members of the CBM35 family. We have expressed a construction containing this domain, and the resulting purified protein has been crystallized and used for soaking experiments to investigate its ligand binding ability. Thus, several crystals were soaked with either GlcA, xylose, xylotriose, xylohexaose, or glucuronic acid-decorated xylooligosaccharides (aldouronic mixture). Only glucuronic acid-containing complexes were obtained. Fig. 5a

Structure of Bimodular *P. barcinonensis* Xyn30D

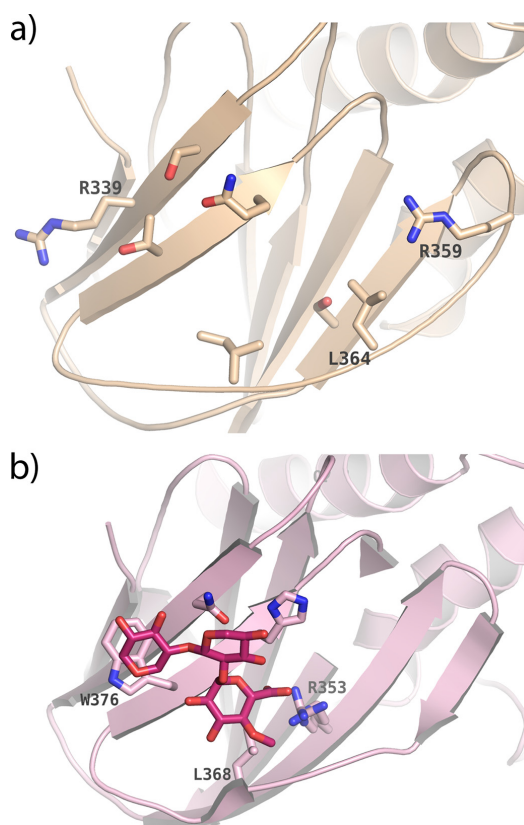


FIGURE 4. The Xyn30D β -fused structure. *a*, a view of Xyn30D β -sandwich showing two arginine residues that could potentially bind the MeGA carboxylate. No aromatic residue is found that could target a xylose moiety. *b*, MeGX2 bound to the BsXynC β -structure. None of the residues involved in binding are conserved among GH30s.

shows the overall fold of the Xyn30D-CBM35 domain and the two observed GlcA binding sites. Site 1 (Fig. 5b) is calcium-dependent and has been observed in the crystals soaked with both GlcA and aldouronic mixture, whereas site 2 has been observed only in the GlcA-soaked crystal. There is a second calcium ion common to many lectins and CBM families having a structural role (Fig. 5d). This calcium is coordinated to Glu¹⁸, Glu²⁰, and Asp¹³⁴ side chains and to the carbonyl group of Thr³⁸, Gly⁴¹, and Asp¹³⁴.

As can be observed in Fig. 5b showing binding site 1, the GlcA moiety is stacked to Trp¹⁰², located at the cleft defined by the loops linking the β -sheets at one side of the barrel and is recognized by Xyn30D-CBM35 through an intricate net of polar interactions, some of them mediated by the calcium ion (Fig. 6). First of all, the C6 carboxylate makes a bidentate interaction with Arg⁷⁹ side chain and is also interacting to Ca1, which in turn interacts also with the GlcA O4. The recognition of the GlcA carboxylate must be critical, because soaking of crystals in nondecorated xylooligosaccharides failed in providing observ-

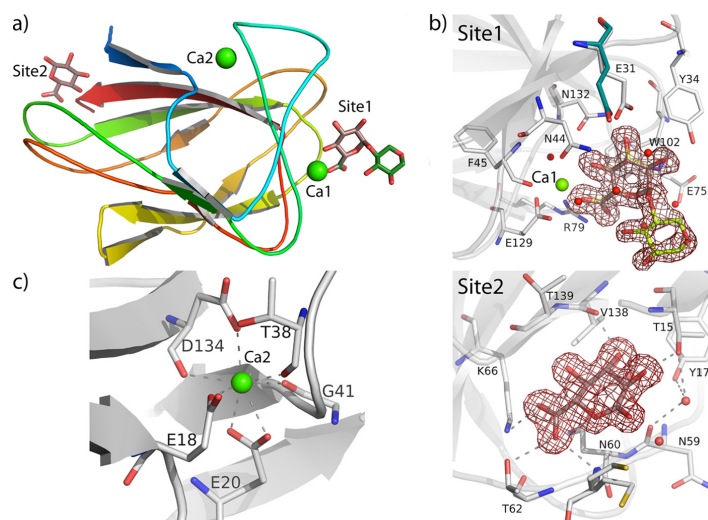
Structure of Bimodular *P. barcinonensis* Xyn30D

FIGURE 5. **The Xyn30D-CBM35.** *a*, cartoon representation of Xyn30D-CBM35 domain colored in rainbow sequence code, showing the two calcium ions as green spheres; the two bound ligands are represented as sticks, the common GlcA moiety being colored in beige, whereas the xylose unit is green. *b*, top panel, view of the calcium-dependent GlcA binding site showing the residues involved as white sticks, and the $2F_o - F_c$ electron density map of the aldouronic-soaked crystal contoured at 1σ . The only change observed upon ligand binding is a rearrangement of Glu³¹ side chain (unbound position shown in slate). Bottom panel, view of the second GlcA molecule found in the GlcA soaked crystal, showing the polar interactions. *c*, the coordination sphere of the structural calcium ion.

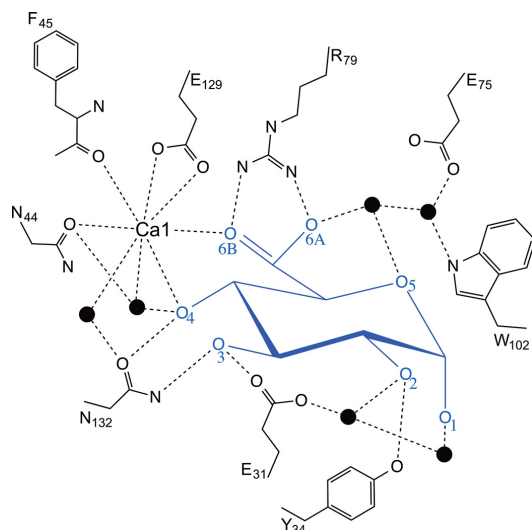


FIGURE 6. **Scheme of the atomic interactions between Xyn30D-CBM35, the calcium ion and the GlcA ligand.**

able complexes. In addition, O1, O2, O3, O4, and O5 also make many polar interactions with the Xyn30D-CBM35 residues Glu³¹, Tyr³⁴, Asn⁴⁴, and Asn¹³², both directly and through several well ordered water molecules, which keeps the GlcA moiety in a very fix position common to the GlcA and aldouronic soaked crystals. On the contrary, the attached xylose unit does not make any direct interaction with residues from the Xyn30D-CBM35 domain, and, in agreement with this observa-

tion, the electron density map shows weaker density at this position precluding building additional xylose units. Finally, the only significant difference observed in the ligand-bound complexes with respect to the unbound structure is a conformational change in the Glu³¹ side chain that rearranges to make a hydrogen bond to GlcA O3 (Fig. 5*b*).

The second GlcA molecule is bound at the loops linking the β -sheets at the opposite side of the β -sandwich, next to the linker connecting the Xyn30D-CBM35 to the catalytic domain (Fig. 5*b*, Site2). The GlcA moiety, clearly seen in the electron density map, makes polar interactions to Val¹³⁸ main chain and to the side chain of Thr¹⁵, through its O2 and O3 hydroxyls. Moreover, the carboxylate at C6 is interacting to Thr⁶² and Lys⁶⁶. However, none of these residues are conserved among CBM35 family and, consequently, binding of GlcA at this site may not have biological significance.

Xyn30D-CBM35 Shows Distinctive Structural Features Regulating Its Specificity—The closest relatives to Xyn30D-CBM35 are the uronic acid-specific CBM35s, which bind Δ 4,5-anhydrogalacturonic acid released from pectin by the action of pectate lyases therefore being a signature of plant cell wall degradation (20). Thus, Xyn30D-CBM35 shows the highest sequence identity with the CBM35 moieties of the exo- β -D-glucosaminidase from *A. orientalis*, Chi-CBM35 (41%), the rhamnogalacturonan acetyl esterase from *Clostridium thermocellum*, Rhe-CBM35 (38%), and the xylanase Xyn10B from *Cellvibrio japonicus*, Xyl-CBM35 (36%), the three being able to bind also GlcA, although Rhe-CBM35 only weakly. The sequence identity is lower (23%) with the other characterized uronic acid-specific CBM35, a pectate lyase from an environmental isolate, Pel-CBM35, which does not show GlcA binding ability. Recognition of uronic acid by these CBM35 domains is always

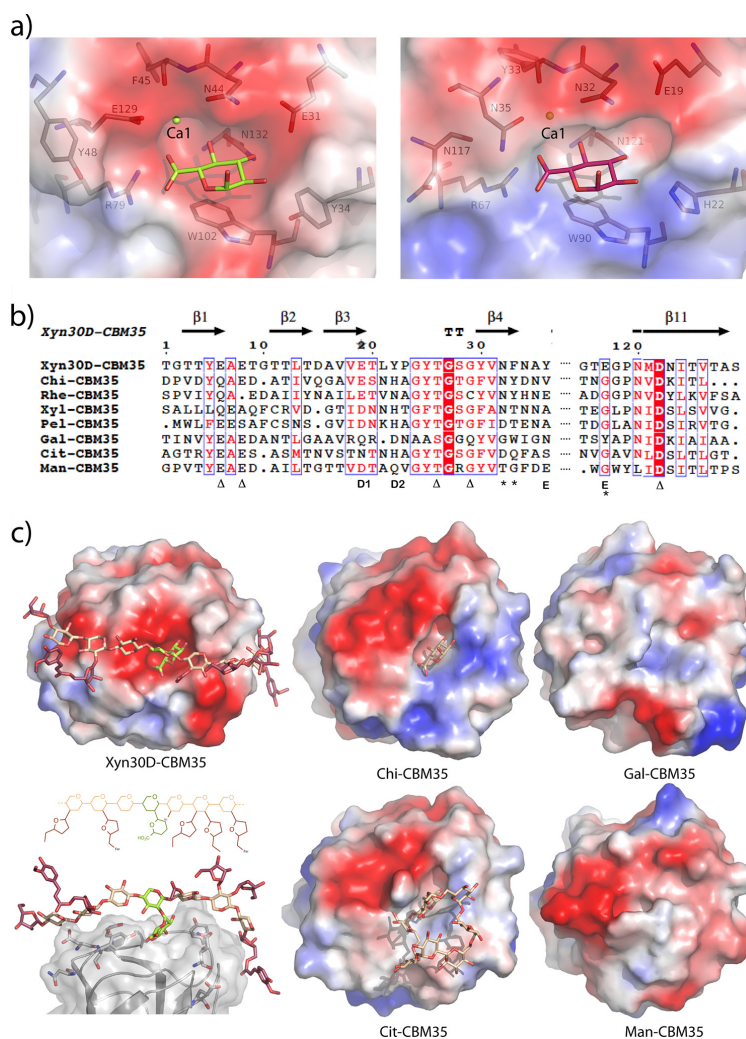
Structure of Bimodular *P. barcinonensis* Xyn30D

FIGURE 7. Structural determinants of Xyn30D-CBM35 specificity. *a*, comparison of the calcium-dependent GlcA binding site (site 1) in Xyn30D-CBM35 (left panel) with its closest homologue, the CBM35 domain of the exo- β -D-glucosaminidase from *A. orientalis*, Chi-CBM35 (right panel). *b*, structural alignment of Xyn30D-CBM35 (DALI) with the structurally known CBM35s: Chi-CBM35 (PDB code 2VZP); rhamnogalacturonan acetyl esterase from *C. thermocellum*, Rhe-CBM35 (PDB code 2W47); xylanase Xyn10B from *C. japonicus*, Xyl-CBM35 (PDB code 2W87); pectate lyase from an environmental isolate, Pel-CBM35 (PDB code 2W3J); CBM35 binding D-galactose from *C. thermocellum*, Gal-CBM35 (PDB code 2WZ8); cycloisomaltooligosaccharide glucanotransferase from *Bacillus circulans*, Cit-CBM35 (PDB code 3WNO); and endo- β -1,4-mannanase from *Podospora anserina*, Man-CBM35 (PDB code 3ZM8). The residues of Xyn30D involved in the functional Ca1 binding are marked with asterisks, whereas residues coordinating the structural Ca2 are indicated with a triangle. Residues located in regions D and E within the binding pocket are involved in modulating ligand specificity within CBM35 domains. *c*, molecular surface of Xyn30D-CBM35 and representative CBM35 domains, showing the charged regions in red (negative) and blue (positive). The GlcA-Xyl ligand found in Xyn30D-CBM35 soaked in aldouronic mixture is shown as green sticks, whereas GlcA and cycloisomaltooligosaccharide found in the Chi-CBM35 and Cit-CBM35 crystals are shown as beige sticks. A model of a glucuronoxarabinosyl chain, represented as sticks, has been superimposed to the GlcA-Xyl ligand found in Xyn30D-CBM35 as explained under "Experimental Procedures."

dependent on calcium, which interacts with the carboxylate of the ligand. However, Xyn30D-CBM35 has a modified version of the (Asp/Asn)-(Tyr/Thr)-Xaa-Asn motif located at the end of β -strand 4 that is conserved in the previously reported uronic acid-CBM35s and builds the metal site (Fig. 7, *a* and *b*). Thus, Asn⁴⁴ and Phe⁴⁵ bind the Ca ion through their side and main

chain, respectively, in the same way observed previously (Asn³²/Tyr³³ in Chi-CBM35), but the terminal Asn³⁵ is missing in the Xyn30D-CBM35 motif, and Glu¹²⁹ from the loop linking β 10- β 11 is occupying an equivalent position to make a bidentate link to Ca1. This Glu¹²⁹ is a determinant of substrate specificity, as will be explained below.

Structure of Bimodular *P. barcinonensis* Xyn30D

Topology seems to be rather conserved through the CBM35 family, as seen by the structural similarities revealed by the DALI server (48). Thus, the RMS values are between 1.0 and 1.8 for 122–127 aligned α atoms of the up to now structurally characterized CBM35 members, given in Fig. 7*b*. Moreover, five regions (A–E) have been identified in the binding site of CBM35s (45) and the related CBM6s (49) as key features determining ligand binding and specificity. Region A and C, highly conserved, are constituted by a Trp, located at the end of β 8 and making stacking interaction to the pyranose ring, and an Asn residue from β 11 interacting with several of its OH groups. These regions are represented by Trp¹⁰² (region A) and Asn¹³² (region C) in the Xyn30D-CBM35 (Fig. 7*a*). Region B is highly variable depending on the subfamily, being the calcium site in the uronate CBM35 binders. Region D is defined by a position occupied by rather conserved polar residues located at β 3, Glu³¹ in Xyn30D-CBM35 that, as explained before, change its position upon GlcA binding. In the uronate-specific CBM35, this region is expanded by an additional and conserved His that interacts to the GlcA O2 hydroxyl, but this His is replaced by Tyr³⁴ in Xyn30D-CBM35 (Fig. 7, *a* and *b*). Finally, region E displays the largest degree of variation and has been stated to play an essential role in ligand specificity presented among CBM35s and CBM6s. This region, defined by the loop before β 11, allocates aromatic residues in some of the domains (Fig. 7*b*); thus, a Tyr in Gal-CBM35 has been described to block off the binding site cleft conferring specificity for the terminal unit of complex carbohydrates, whereas two Trps in Man-CBM35 have been attributed a role extending the binding cleft to additional subsites able to bind mano-oligosaccharides (45). In contrast to the known uronate-binding CBM35 that does not exhibit particular features at this region, Xyn30-CBM35 contains Glu¹²⁹, which is essential in calcium binding, as said before. Furthermore, and close to this position, Tyr⁴⁸ from loop β 4– β 5 is protruding at the binding cleft pointing to a putative role in ligand recognition (Fig. 7, *a* and *b*). Consequently, a distinctive feature of Xyn30D-CBM35 with respect to the other uronate-binding CBM35s is the presence of two nonconserved aromatic residues delineating the binding pocket, *i.e.* Tyr³⁴ and Tyr⁴⁸, which is an interesting trait considering the involvement of Tyr in carbohydrate recognition that therefore might result in additional binding subsites.

Fig. 7*c* displays the molecular envelop of Xyn35D-CBM35, together with that of Chi-CBM35, as a representative member of the uronate binding subgroup and the other structurally known CBM35 domains. As can be observed in the figure, the sequence variability among the family allows for a great structural diversity of the surface shape that must count for the diverse ligands recognized by each domain. Remarkably, Xyn30D-CBM35 shows a more open pocket, as compared with the uronate-binding Chi-CBM35. Also, it is unique in presenting a continuous stretch of negative charge distribution extending from its binding pocket, which could point to some sort of oriented interaction with its target substrate. A xylooligosaccharide chain with a decoration pattern typical of glucuronarabinoxylans found in grass cell walls (5) has been modeled as explained under “Experimental Procedures.” This chain has been docked onto the active site surface, by superposition of

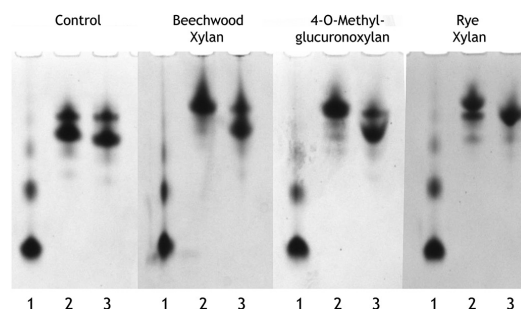


FIGURE 8. Affinity gel electrophoresis analysis of Xyn-CBM35 binding to soluble xylans. Xyn-CBM35 was analyzed on nondenaturing polyacrylamide gels containing no ligand (control) or soluble xylan from beechwood, hardwoods (4-O-methyl glucuronoxylan), or rye. Lanes 1, BSA; lanes 2, Xyn-CBM35; lanes 3, Xyn-CBM35, 15 mM EDTA.

its GlcA-xylose moiety onto the experimental position of this disaccharide found in the Xyn30D-CBM35 soaked crystals. As can be observed in Fig. 7*c*, the modeled conformation, which keeps the typical 3-fold shape of xylan chain, appears complementary to the binding site topology that, on the other hand, seems able to allocate also Ara and GlcA decorations at O2/O3 in all the xylose units. This surface topology would justify the previously reported binding properties of Xyn30D-CBM35 to different types of decorated xylans (27). Moreover, an inspection to the Xyn30D-CBM35 structure reveals that, apart from the two mentioned Tyr³⁴ and Tyr⁴⁸, no other aromatic residues are evident at the binding site surface that might shape typical substrate binding sites. However, many polar residues might be making stabilizing interactions with the putative substrate (Fig. 7*c*). Consequently, although the Xyn30D-CBM35 domain is a GlcA-specific binder, other preferred interactions with its substrate increasing affinity cannot be excluded. To evaluate the importance of GlcA for interactions with substrate and their calcium dependence, the binding ability of Xyn30D-CBM35 in the presence of calcium chelating EDTA was analyzed by affinity gel electrophoresis. Migration of the purified module in gels containing soluble xylans was retarded when compared with the migration of the protein in gels without xylan (Fig. 8). Both glucuronoxylan, from beechwood and hardwoods, and glucuronarabinoxylan from rye produced a retardation of the CBM35. However, retardation was significantly inhibited when EDTA was included in gels, clearly showing the involvement of the calcium-dependent GlcA binding, as indicated above. Notably, although retardation by glucuronoxylans was almost completely reverted by EDTA, retardation by rye glucuronarabinoxylan was less affected by the chelating agent. This would indicate that in addition to the GlcA interaction, the Xyn30D-CBM35 might establish interaction to other components of xylan molecule, such as arabinose decorations of glucuronarabinoxylan, as suggested.

DISCUSSION

The biological conversion of the polysaccharides within the plant cell wall to their constituent monosaccharides is performed by a large number of enzymes that are as diverse as their complex substrates. An example of this chemical complexity is

| | |
|--------------|----------------|
| H6WCZ0_PAEB | LSGGNSGGNVN |
| O24852_AERCA | LTGGN-GGGNVG |
| W4C6X9_9BACL | VTGGN-SGGNTG |
| P70733_AERCA | ITGGS---VTPG |
| G7W2J1_PAETH | LGG-----KTI |
| E0RDU2_PAEF6 | LGGKPI SN---- |
| H6CPJ0_9BACL | LGKNSNIV-NTN |
| B8I0M3_CLOCE | LGTITPVEKDAF |
| U4QXW6_9CLOT | LGTITPVEKDAF |
| E6UTM3_CLOTL | GSSEPQIPVERNAF |

FIGURE 9. Linker sequence of the closest modular Xyn30D analogues. Six sequences present the domain composition GH30-CBM35, whereas eleven more are GH30-CBM6 (three representative members are given).

provided by xylan, the major hemicellulosic component of the cell wall. The great majority of xylanases coping with this diversity belong to GH10 and GH11, a variety of debranching enzymes being associated to them to accomplish efficiently the degradation process. More recently, new endoxylanase activities have been identified in families GH5 and GH30 that, unlike classical xylanases, are able to function in high substituted xylans and, even more, they use decorations as specificity determinants. Xyn30D is one of the glucuronoxylan-specific endoxylanases included in GH30.

Within the complex landscape of plant cell wall degrading process, modularity is a common feature used by the implicated enzymes to fine-tune their biological functionality through attached CBMs. It is generally observed that the linker regions between catalytic and CBM display a great deal of structural flexibility to maximize substrate accessibility. According to this flexibility, only a few bacterial modular enzymes have been structurally characterized up to now. With a few exceptions (17, 35), the isolated modules can generally function independently, and this is the case for Xyn30D (27). In despite of this, we have been able to crystallize the full-length enzyme, and this is the first intact structure that allows visualization of the juxtaposition of the CBM35 module relative to the GH30 catalytic domain.

A blast search gives ~100 entries showing more than 50% identity with Xyn30D sequence. Six of these sequences include a CBM35 domain attached to the catalytic GH30 domain and represent hydrolytic enzymes from different *Paenibacillus* and *Aeromonas* species. Alignment of the corresponding sequences reveals that they all display a short linker region of 6–12 Gly/Pro-rich residues (Fig. 9). Consequently, the overall molecular shape of Xyn30D here presented should be a good model for these enzymes. Moreover, 11 more sequences correspond to modular enzymes including, instead of CBM35, a topologically related CBM6 attached to the C terminus of the chain. These sequences correspond to different enzymes from *Clostridium*, all of them having also an additional C-terminal docking domain. Although the sequence is less conserved in the linker region (Fig. 9), the similar size suggests that the arrangement of the catalytic and the CBM6 domains might be not too different to that observed in Xyn30D.

A different situation is represented by the arabinoxylan-specific xylanase Xyl5A from *Clostridium thermocellum* (CtXyl5A),

Structure of Bimodular *P. barcinonensis* Xyn30D

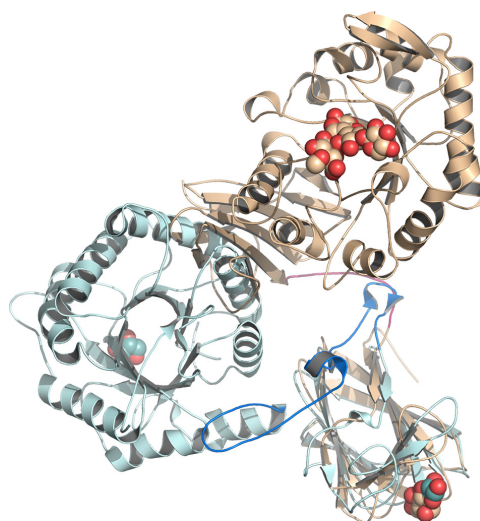


FIGURE 10. The molecular architecture of Xyn30D versus CtXyl5A. The CBM35 moiety of Xyn30D (beige) has been superimposed to the CBM6 module of CtXyl5A (cyan). Ligands showing the catalytic and their common CBM binding sites are shown as spheres. The linkers are colored pink (Xyn30D) and blue (CtXyl5A).

a bimodular xylanase with a domain structure GH5-CBM6 closely related to Xyn30D (17). Similarly to Xyn30D, CtXyl5A uses xylan decorations as determinants of specificity binding arabinose at its subsite –1 within the active site. Moreover, its ancillary CBM6 domain recognizes the termini of xylo- and gluco-configured polymers and likewise Xyn30D-CBM35 homologues; it has been proposed to target the enzyme to regions of the plant cell wall that are being degraded (17). However, the superposition of Xyl5A to Xyn30D (Fig. 10) shows a very different molecular architecture in which the catalytic and the CBM domains are oppositely oriented through their distinct linker regions. This picture highlights the diverse structural scaffolds generated by the combination of, in principle, structurally related domains, which must result in a great functional plasticity. Furthermore, it emphasizes the importance of studying entire modular enzymes for a comprehensive understanding of the molecular machinery involved in the degradation of the complex plant cell wall.

Members of the genus *Bacillus* are common saprophytic components of soil microbiota. *P. barcinonensis* sp., isolated from rice fields, has been reported to be a high level producer of xylanases in media supplemented with rice straw, demonstrating the presence of a complex enzymatic system for xylan degradation. In particular Xyn30D, in agreement with other reported GH30 enzymes, shows high activity on glucuronoxylans but none on arabinoxylans. Recognition of a methylglucuronic substitution is required for hydrolysis of the xylan chain. On the other hand, its attached CBM35 specifically binds non-methylated glucuronic acid decorations of different xylans in a calcium-dependent manner. The binding is more efficient with glucuronoxylans and glucuronoarabinoxylans, the last being probably the more common xylan type in the natural habitat of

Structure of Bimodular *P. barcinonensis* Xyn30D

Xyn30D. Our results suggest the ability of Xyn30D-CBM35 to accommodate substitutions in all the xylose units, which enables this module to bind highly decorated xylans. Furthermore, we found that, apart from its GlcA target, it might establish interaction to other components of xylan molecule, such as arabinose decorations of glucuronoarabinoxylan. The concerted action of the different domains is intriguing and is difficult to envisage the precise role that Xyn30D play in xylan degradation, but its peculiar structural features point to a role in depolymerization of highly substituted, chemically complex xylans. The methylated/unmethylated substrate specificity of its catalytic and ancillary domains probably reflects a sophisticated enzyme delivery strategy of *Paenibacillus* related to the complexity of its natural substrate.

More work has to be done to increase our knowledge on the molecular mechanisms regulating modularity. Apart from its biological interest, it will help in producing new, more efficient biocatalysts for the conversion of lignocelluloses to value-added products and fuels. It will result in more environmentally sustainable industries.

Acknowledgments—We thank the staff of the European Synchrotron Radiation Facility (Grenoble, France) for providing access and for technical assistance at Beamline ID23.1. We also thank the German Electron Synchrotron (Hamburg, Germany) for assistance at Petra III Beamline and the Synchrotron Radiation Source (Daresbury, UK) for assistance at I03 Beamline.

REFERENCES

- Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., Frederick, W. J., Jr., Hallett, J. P., Leak, D. J., Liotta, C. L., Mielenz, J. R., Murphy, R., Templer, R., and Tschaplinski, T. (2006) The path forward for biofuels and biomaterials. *Science* **311**, 484–489
- Tuck, C. O., Pérez, E., Horváth, I. T., Sheldon, R. A., and Poliakoff, M. (2012) Valorization of biomass: deriving more value from waste. *Science* **337**, 695–699
- Collins, T., Gerday, C., and Feller, G. (2005) Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol. Rev.* **29**, 3–23
- de Vries, R. P., and Visser, J. (2001) Aspergillus enzymes involved in degradation of plant cell wall polysaccharides. *Microbiol. Mol. Biol. Rev.* **65**, 497–522
- Scheller, H. V., and Ulvskov, P. (2010) Hemicelluloses. *Annu. Rev. Plant Biol.* **61**, 263–289
- Grootaert, C., Van den Abbeele, P., Marzorati, M., Broekaert, W. F., Courtin, C. M., Delcour, J. A., Verstraete, W., and Van de Wiele, T. (2009) Comparison of prebiotic effects of arabinoxylan oligosaccharides and inulina simulator of the human intestinal microbial ecosystem. *FEMS Microbiol. Ecol.* **69**, 231–242
- Jang, Y. S., Kim, B., Shin, J. H., Choi, Y. J., Choi, S., Song, C. W., Lee, J., Park, H. G., and Lee, S. Y. (2012) Bio-based production of C2-C6 platform chemicals. *Biotechnol. Bioeng.* **109**, 2437–2459
- Meng, X., and Ragauskas, A. J. (2014) Recent advances in understanding the role of cellulose accessibility in enzymatic hydrolysis of lignocellulosic substrates. *Curr. Opin. Biotechnol.* **27**, 150–158
- Leu, S.-Y., and Zhu, J. Y. (2013) Substrate-related factors affecting enzymatic saccharification of lignocelluloses: our recent understanding. *Bioenerg. Res.* **6**, 405–415
- Wyman, C. E., Balan, V., Dale, B. E., Elander, R. T., Falls, M., Hames, B., Holtzapfle, M. T., Ladisch, M. R., Lee, Y. Y., Mosier, N., Pallapolu, V. R., Shi, J., Thomas, S. R., and Warner, R. E. (2011) Comparative data on effects of leading pretreatments and enzyme loadings and formulations on sugar yields from different switchgrass sources. *Bioresour. Technol.* **102**, 11052–11062
- Pollet, A., Delcour, J. A., and Courtin, C. M. (2010) Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families. *Crit. Rev. Biotechnol.* **30**, 176–191
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495
- Kolenová, K., Vrsanská, M., and Biely, P. (2006) Mode of action of endo- β -1,4-xylanases of families 10 and 11 on acidic xylooligosaccharides. *J. Biotechnol.* **121**, 338–345
- Gallardo, O., Fernández-Fernández, M., Valls, C., Valenzuela, S. V., Roncero, M. B., Vidal, T., Díaz, P., and Pastor, F. I. (2010) Characterization of a family GH5 xylanase with activity on neutral oligosaccharides and evaluation as a pulp bleaching aid. *Appl. Environ. Microbiol.* **76**, 6290–6294
- Vrsanská, M., Kolenová, K., Puchart, V., and Biely, P. (2007) Mode of action of glycoside hydrolase family 5 glucuronoxylan xylanohydrolase from *Erwinia chrysanthemi*. *FEBS J.* **274**, 1666–1677
- St John, F. J., Rice, J. D., and Preston, J. F. (2006) Characterization of XynC from *Bacillus subtilis* subsp. *subtilis* strain 168 and analysis of its role in depolymerization of glucuronoxylan. *J. Bacteriol.* **188**, 8617–8626
- Correia, M. A., Mazumder, K., Brás, J. L., Firbank, S. J., Zhu, Y., Lewis, R. J., York, W. S., Fontes, C. M., and Gilbert, H. J. (2011) Structure and function of an arabinoxylan-specific xylanase. *J. Biol. Chem.* **286**, 22510–22520
- Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* **382**, 769–781
- Gilbert, H. J., Knox, J. P., and Boraston, A. B. (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Curr. Opin. Struct. Biol.* **23**, 669–677
- Montanier, C., van Bueren, A. L., Dumon, C., Flint, J. E., Correia, M. A., Prates, J. A., Firbank, S. J., Lewis, R. J., Grondin, G. G., Ghinet, M. G., Gloster, T. M., Herve, C., Knox, J. P., Talbot, B. G., Turkenburg, J. P., Kerovuo, J., Brzezinski, R., Fontes, C. M., Davies, G. J., Boraston, A. B., and Gilbert, H. J. (2009) Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3065–3070
- Guillén, D., Sánchez, S., and Rodríguez-Sanoja, R. (2010) Carbohydrate-binding domains: multiplicity of biological roles. *Appl. Microbiol. Biotechnol.* **85**, 1241–1249
- Shoseyov, O., Shani, Z., and Levy, I. (2006) Carbohydrate binding modules: biochemical properties and novel applications. *Microbiol. Mol. Biol. Rev.* **70**, 283–295
- Sánchez, M. M., Fritze, D., Blanco, A., Spröer, C., Tindall, B. J., Schumann, P., Kroppenstedt, R. M., Díaz, P., and Pastor, F. I. (2005) *Paenibacillus barcinonensis* sp. nov., a xylanase-producing bacterium isolated from a rice field in the Ebro River delta. *Int. J. Syst. Evol. Microbiol.* **55**, 935–939
- Valenzuela, S. V., Díaz, P., and Pastor, F. I. (2014) Xyn11E from *Paenibacillus barcinonensis* BP-23: a LppX-chaperone-dependent xylanase with potential for upgrading paper pulps. *Appl. Microbiol. Biotechnol.* **98**, 5949–5957
- Valenzuela, S. V., Díaz, P., and Javier Pastor, F. I. (2010) Recombinant expression of an alkali stable GH10 xylanase from *Paenibacillus barcinonensis*. *J. Agric. Food. Chem.* **58**, 4814–4818
- Gallardo, O., Pastor, F. I., Polaina, J., Díaz, P., Lysek, R., Vogel, P., Isorna, P., González, B., and Sanz-Aparicio, J. (2010) Structural insights into the specificity of Xyn10B from *Paenibacillus barcinonensis* and its improved stability by forced protein evolution. *J. Biol. Chem.* **285**, 2721–2733
- Valenzuela, S. V., Díaz, P., and Pastor, F. I. (2012) Modular glucuronoxylan-specific xylanase with a family CBM35 carbohydrate-binding module. *Appl. Environ. Microbiol.* **78**, 3923–3931
- St John, F. J., Hurlbert, J. C., Rice, J. D., Preston, J. F., and Pozharski, E. (2011) Ligand bound structures of a glycosyl hydrolase family 30 glucuronoxylan xylanohydrolase. *J. Mol. Biol.* **407**, 92–109
- Urbániková, L., Vrsanská, M., Mørkeberg Krogh, K. B., Hoff, T., and Biely, P. (2011) Structural basis for substrate recognition by *Erwinia chrysanthemi* GH30 glucuronoxylanase. *FEBS J.* **278**, 2105–2116
- Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685

Structure of Bimodular *P. barcinonensis* Xyn30D

31. Sainz-Polo, M. Á., Valenzuela, S. V., Pastor, F. J., and Sanz-Aparicio, J. (2014) Crystallization and preliminary x-ray diffraction analysis of Xyn30D from *Paenibacillus barcinonensis*. *Acta Crystallogr. F Struct. Biol. Commun.* **70**, 963–966
32. Batty, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D* **67**, 271–281
33. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242
34. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
35. Couturier, M., Roussel, A., Rosengren, A., Leone, P., Stålbrand, H., and Berrin, J. G. (2013) Structural and biochemical analyses of glycoside hydrolase families 5 and 26 β -(1,4)-mannanases from *Podospora anserina* reveal differences upon manno-oligosaccharide catalysis. *J. Biol. Chem.* **288**, 14624–14635
36. Vagin, A. A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025
37. Stein, N. (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.* **41**, 641–643
38. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255
39. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132
40. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291
41. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21
42. DeLano, W. L. (2002) PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA
43. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* **29**, 622–655
44. Leggio, L. L., Jenkins, J., Harris, G. W., and Pickersgill, R. W. (2000) X-ray crystallographic study of xylopentaose binding to *Pseudomonas fluorescens* xylanase A. *Proteins* **41**, 362–373
45. Correia, M. A., Abbott, D. W., Gloster, T. M., Fernandes, V. O., Prates, J. A., Montanier, C., Dumon, C., Williamson, M. P., Tunncliffe, R. B., Liu, Z., Flint, J. E., Davies, G. J., Henrissat, B., Coutinho, P. M., Fontes, C. M., and Gilbert, H. J. (2010) Signature active site architectures illuminate the molecular basis for ligand specificity in family 35 carbohydrate binding module. *Biochemistry* **49**, 6193–6205
46. St John, F. J., González, J. M., and Pozharski, E. (2010) Consolidation of glycosyl hydrolase family 30: a dual domain 4/7 hydrolase family consisting of two structurally distinct groups. *FEBS Lett.* **584**, 4435–4441
47. Larson, S. B., Day, J., Barba de la Rosa, A. P., Keen, N. T., and McPherson, A. (2003) First crystallographic structure of a xylanase from glycoside hydrolase family 5: implications for catalysis. *Biochemistry* **42**, 8411–8422
48. Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549
49. Abbott, D. W., Ficko-Blean, E., van Bueren, A. L., Rogowski, A., Cartmell, A., Coutinho, P. M., Henrissat, B., Gilbert, H. J., and Boraston, A. B. (2009) Analysis of the structural and functional diversity of plant cell wall specific family 6 carbohydrate binding modules. *Biochemistry* **48**, 10395–10404

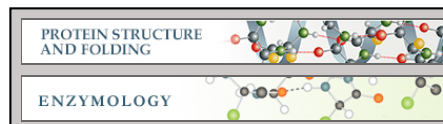


Protein Structure and Folding:
Structural Analysis of
Glucuronoxylan-specific Xyn30D and Its
Attached CBM35 Domain Gives Insights
into the Role of Modularity in Specificity

M. Angela Sainz-Polo, Susana Valeria
 Valenzuela, Beatriz González, F. I. Javier
 Pastor and Julia Sanz-Aparicio

J. Biol. Chem. 2014, 289:31088-31101.

doi: 10.1074/jbc.M114.597732 originally published online September 8, 2014



Access the most updated version of this article at doi: [10.1074/jbc.M114.597732](https://doi.org/10.1074/jbc.M114.597732)

Find articles, minireviews, Reflections and Classics on similar topics on the [JBC Affinity Sites](#).

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 48 references, 16 of which can be accessed free at
<http://www.jbc.org/content/289/45/31088.full.html#ref-list-1>

4.4.3. Resultados

En este trabajo se presenta la estructura tridimensional de la proteína completa Xyn30D de *P. barcinonensis*, una xilanasa modular con una composición de dominios GH30-CBM35 que ha sido resuelta a 2,4 Å de resolución. El grupo espacial resultó ser $P3_2$, con ocho moléculas en la unidad asimétrica. Así mismo, se describe la estructura de los complejos del dominio no catalítico, que denominamos Xyn30D-CBM35, con ácidos glucurónico y aldurónico. Además, se han realizado ensayos de unión de este dominio con distintos xilanos mediante geles nativos de afinidad.

El estudio cristalográfico de Xyn30D muestra que el dominio catalítico tiene un plegamiento típico de la familia GH30 formado a su vez por dos regiones: un barril $(\beta/\alpha)_8$ y una subestructura tipo sándwich β , localizándose el sitio catalítico en el eje del barril. Además, y unido a su extremo C-terminal, se encuentra el dominio CBM35, con una topología de láminas β tipo lectina, que está dispuesto de forma casi perpendicular y orientado hacia el lado opuesto al centro catalítico. Dentro del mismo se ubican dos iones calcio, Ca1 y Ca2, con papel funcional y estructural, respectivamente. Ambos dominios se pliegan de forma independiente y sólo se observan pequeñas diferencias en la orientación de uno respecto del otro dentro de las ocho moléculas que forman la unidad asimétrica. Aún así, se debe tener en cuenta que parte del segmento conector está estabilizado mediante la formación de enlaces de hidrógeno con residuos del módulo catalítico, por lo que en ausencia de restricciones relacionadas con el empaquetamiento cristalino, no se puede descartar una mayor flexibilidad molecular.

Respecto al análisis del dominio catalítico Xyn30D-CD, cabe resaltar que su subestructura β no presenta sitio de unión para MeGlcA, por lo que no puede desempeñar el papel de unión a sustrato observada en otras enzimas de la familia, como por ejemplo en la xilanasa de *Bacillus* (BsXyn10C). Por consiguiente, puede descartarse que este sea el papel general del subdominio en todos los miembros de la familia GH30, al menos en las enzimas GH30 modulares.

Se realizaron muchos intentos para obtener cristales de Xyn30D con sus sustratos, todos fueron infructuosos. Analizando su estructura, se comprobó que el segmento conector y el extremo N-terminal que contiene la cola de histidinas se encuentran ocluyendo el sitio activo de una molécula contigua relacionada por simetría. Por consiguiente, la difusión de los ligandos al sitio de unión probablemente alteraría esta interacción, la cual debe ser importante para la estabilización del cristal. En un intento por superar este problema, se realizaron construcciones de los módulos catalítico y no catalítico de forma independiente. Sólo Xyn30D-CBM35 cristalizó y se abordaron los experimentos de cristalización frente a varios posibles sustratos.

Hay que destacar que únicamente se obtuvieron los complejos de Xyn30D-CBM35 con sustratos que contenían GlcA, lo que apunta a que este es un determinante específico de unión de Xyn30D-CBM35. Se distinguen en su estructura dos sitios de unión. Uno de ellos, que es dependiente de calcio, se observa en los cristales tratados con GlcA y ácido aldurónico. Una diferencia significativa observada en los complejos con respecto a la enzima sin sustrato es el cambio conformacional que presenta la cadena lateral del residuo Glu31, la cual se reorganiza para formar un enlace de hidrógeno con el O3 de GlcA. El segundo sitio únicamente se observa cuando se sumergen los cristales en GlcA y ninguno de los residuos implicados en esta unión está conservado dentro de la familia CBM35. Así, en consecuencia, la unión de GlcA en este segundo sitio no parece tener relevancia biológica.

Se han descrito previamente cinco regiones (A-E) en el sitio de unión de la familia CBM35 como distintivos esenciales que determinan la unión del ligando. Xyn30D-CBM35 muestra características estructurales propias en estas regiones que regulan su especificidad. Por ejemplo, el Glu129 no descrito en la familia CBM35 como determinante en la unión, presenta un enlace bidentado con el calcio, por lo que parece esencial. Este residuo forma parte de la región E, la cual ha sido considerada como *hotspot* para la especificidad de sustrato dentro de esta familia. Además, cerca de esta posición sobresalen dos tirosinas de la hendidura de unión, Tyr34 y Tyr48, las cuales apuntan a un posible papel en el reconocimiento del ligando. Estos dos residuos aromáticos no conservados representarían un rasgo particular de Xyn30D-CBM35 y podría sugerir la existencia de subsitios de unión adicionales.

Se evaluó la importancia del calcio en la unión de Xyn30D-CBM35 al sustrato mediante geles de afinidad en presencia y ausencia de EDTA, lo que permitió concluir que la unión es claramente dependiente de calcio. Un resultado novedoso obtenido de estos ensayos fue que la unión de Xyn30D-CBM35 a GAX no se vio afectada del mismo modo que otros sustratos por la eliminación del calcio. Esto sugiere la posible existencia de otras interacciones con GAX en sitios distintos al de reconocimiento de GlcA dependiente de calcio.

La comparación de la superficie molecular de Xyn30D-CBM35 con miembros representativos de la familia CBM35 muestra características importantes relativas a su función. La variabilidad de secuencia dentro de esta familia permite una gran diversidad estructural, de forma que la superficie de estos dominios es muy variable según el ligando que reconozcan. Sorprendentemente, Xyn30D-CBM35 presenta una hendidura de unión muy abierta con una distribución de carga negativa continua que se extiende a lo largo del sitio de enlace. Esto puede apuntar a algún tipo de interacción orientada con su sustrato diana, adicional al sitio esperado de reconocimiento de GlcA, tal y como sugerían los ensayos de afinidad en presencia de EDTA. Para poder establecer cómo podría ser el reconocimiento, se ha modelado un xilooligosacárido con un patrón de sustitución típico de GAX que se encuentra en las paredes celulares de plantas herbáceas. Para ello, se ha partido de los

datos experimentales obtenidos del complejo de Xyn30D-CBM35 con ácido aldurónico. Los resultados muestran una extraordinaria complementariedad del sustrato a la topología del sitio de unión que, por otro lado, parece capaz de albergar también sustituciones de arabinosa y GlcA en las posiciones O2/O3 de todas las moléculas de xilosa.

4.4.4. Discusión

Es bien conocido que se requiere de la acción coordinada de una gran cantidad de enzimas de diferentes familias para una biodegradación completa y eficaz del xilano. Frente a las enzimas de las familias GH10 y GH11, que han sido las mejor caracterizadas hasta el momento, se están identificando un número creciente de nuevas familias con actividad endoxilanasas. Un ejemplo serían las enzimas de las familias GH5 y GH30, las cuales requieren de “decoraciones”, ramificaciones específicas en la cadena de xilosas, para su hidrólisis. La enzima de estudio del presente trabajo, Xyn30D, es una endoxilanasas específica de GX incluida en la familia GH30.

El complejo entramado que forma la pared celular vegetal hace que su degradación sea un proceso muy complicado debido a la dificultad de acceso a la misma. En este contexto, la modularidad es una estrategia bastante común adoptada por las enzimas implicadas, ya que pueden dirigir su actividad biológica a través de la combinación de distintas especificidades presentadas por dominios suplementarios al catalítico. La mayor parte de estos dominios son CBM. Generalmente se observa que el segmento que une los dominios catalítico y accesorios suele ser flexible, aumentando así la plasticidad y adaptabilidad de la enzima, de manera que se optimiza la accesibilidad al sustrato altamente heterogéneo. Sin embargo, esta flexibilidad supone una dificultad para el estudio de enzimas modulares por cristalografía, por lo que aún son pocas las xilanasas modulares caracterizadas hasta la fecha. Por otra parte, y dado que los CBM pueden funcionar independientemente, la información de la que se dispone hasta la fecha sobre su actividad y especificidad proviene mayoritariamente de CBM aislados.

En este contexto se abordó el presente trabajo y, a pesar de ser una enzima modular, se logró cristalizar y resolver la estructura de la proteína completa mediante cristalografía de rayos X. Así, es la primera estructura resuelta que permite visualizar la disposición de un CBM35 unido a un GH30. Previos estudios bioquímicos indicaban que Xyn30D de *P. barcinonensis* tenía actividad exclusiva sobre metilglucuronoarabinoxilano (MeGAX), y no sobre AX. El CBM35 incluido en Xyn30D es raramente encontrado en xilanasas lo que hace que la enzima en estudio sea excepcional. Ensayos de ITC mostraron que unía GlcA y requería de calcio para ello. Estos resultados están en consonancia con los de otras enzimas ya caracterizadas (Montanier *et al.*, 2009). En el presente trabajo, la resolución estructural de Xyn30D y Xyn30D-CBM35 pone de manifiesto los determinantes estructurales claves presentes en la unión.

El reconocimiento del carboxilato de GlcA es crucial, ya que la inmersión de los cristales en XOS no sustituidos no condujo a la obtención de complejos. Además, el reconocimiento de GlcA es altamente específico y existen muchas interacciones polares tanto de forma directa como a través de varias moléculas de agua perfectamente ordenadas, que mantienen la molécula de GlcA en una posición fija. A este respecto, se han descubierto nuevos determinantes de unión dentro de la familia CBM35, como sería el Glu129 y los residuos aromáticos Tyr48 y Tyr34, que sugieren la existencia de subsitios de unión adicionales a los ya definidos hasta el momento. Por otra parte, el hecho de que el calcio no sea esencial en la unión a GAX podría indicar que, además de la interacción con GlcA, Xyn30D-CBM35 podría establecer una interacción más inespecífica con otros componentes de la molécula de xilano, tales como sustituciones de arabinosa, que situarían el sustrato de una manera orientada. Por tanto, todo apunta a que Xyn30D-CBM35 es capaz de unir cadenas muy ramificadas de GAX.

Pensamos que Xyn30D podría ser un buen modelo estructural para una serie de enzimas identificadas previamente que presentan similar composición de dominios y con segmentos conectores de longitud y secuencia equivalentes. Hay que destacar el papel clave de estos segmentos conectores como determinantes de la especificidad enzimática. Por ejemplo, se ha descrito la estructura de la xilanasa modular Xyl5A de *Clostridium thermocellum*, con una composición de dominios GH5-CBM6 ambos muy relacionados con los presentes en Xyn30D en cuanto a especificidad. Sin embargo, esta xilanasa presenta un segmento conector muy distinto y una orientación de dominios completamente diferente a la descrita en Xyn30D. Cabe esperar, por tanto, un sitio de unión y una función biológica bien distinta respecto a Xyn30D. Esto refleja la importancia del estudio de las enzimas modulares completas, como el presentado aquí, para poder comprender claramente su mecanismo de acción.

Los resultados sugieren la alta capacidad que presenta Xyn30D-CBM35 de acomodar sustituciones en todas las unidades de xilosa. Aunque es difícil predecir la función exacta que tiene Xyn30D en la degradación de xilano, sus peculiares características estructurales y de especificidad apuntan a un papel clave en la despolimerización de xilanos altamente sustituidos. De este modo, es posible que Xyn30D-CBM35 dirija a la enzima a regiones de la pared celular vegetal que están siendo degradadas, ya que muestra alta especificidad por sustratos derivados de la degradación de la pectina como GlcA. Por su parte, el reconocimiento de sustituciones de metil glucurónico por el dominio catalítico es clave para la hidrólisis de la cadena del polímero. Además, el CBM35 adjunto se une específicamente a sustituyentes de GlcA no metilado, siendo este enlace dependiente de calcio. La unión es más eficiente frente a GXs y GAXs, siendo este último, probablemente, el xilano más común en el hábitat natural de Xyn30D. La distinta especificidad de los dominios catalítico y CBM frente a un sustrato metilado (más voluminoso y menos accesible) y a uno no metilado (menos voluminoso y más accesible) respectivamente, evidencia la sofisticada estrategia de degradación del xilano que desarrolla *Paenibacillus*. Este

complicado mecanismo de degradación es un reflejo de la complejidad que probablemente presenta su sustrato natural.

Se debe seguir investigando este tipo de enzimas modulares para comprender mejor los complejos mecanismos moleculares que regulan la degradación de la pared celular vegetal. A parte del importante interés biológico que presentan, la comprensión de su bioquímica ayudará al diseño y producción de nuevos y más eficientes biocatalizadores para la conversión de lignocelulosas en productos de gran valor añadido. Todo ello permitirá diseñar industrias más sostenibles.

4.4.5. Conclusiones: Artículos 3 y 4

Las conclusiones de los trabajos “Artículo 3” y “Artículo 4”, se recogen a continuación:

1. Se ha diseñado un protocolo de purificación de Xyn30D para producir muestra de calidad adecuada para los experimentos de cristalización y difracción. Se han obtenido cristales de Xyn30D y se ha recogido su espectro de difracción. Se ha determinado su grupo espacial, $P3_2$, así como sus parámetros de celdilla.
2. Se ha resuelto la estructura cristalográfica de la proteína Xyn30D completa a 2,4 Å de resolución. Es la primera proteína bimodular de la familia GH30 cristalizada. Esto ha permitido ver cómo está dispuesto su módulo Xyn30D-CBM35 en relación al dominio catalítico.
3. Se ha diseñado un protocolo de expresión y purificación del dominio no catalítico Xyn30D-CBM35 que permite obtener muestra de calidad.
4. Se han establecido las condiciones de cristalización óptimas para conseguir cristales de Xyn30D-CBM35 de calidad suficiente para los experimentos de difracción. También se ha conseguido obtener cristales de Xyn30D-CBM35 en complejo con GlcA y ácido aldurónico. Se ha determinado el grupo espacial de Xyn30D-CBM35 y de sus complejos Xyn30D-CBM35/GlcA y Xyn30D-CBM35/GlcA-Xyl, así como sus parámetros de celdilla.
5. Se ha resuelto la estructura cristalográfica de Xyn30D-CBM35 y de sus complejos con GlcA y ácido aldurónico, a 1,50 Å y 1,35 Å de resolución, respectivamente.
6. Xyn30D presenta una estructura bimodular. El dominio catalítico presenta un plegamiento tipo barril (β/α)₈, unido a una estructura β tipo sándwich de nueve hebras, exclusiva de esta familia. Xyn30D-CBM35 presenta un plegamiento β tipo lectina, dispuesto de forma casi perpendicular y hacia el lado opuesto al sitio catalítico.
7. A diferencia de otras proteínas modulares, Xyn30D presenta moderada flexibilidad molecular, observándose pocas diferencias conformacionales en las ocho moléculas presentes en la unidad asimétrica en el cristal.
8. El subdominio sándwich β del dominio catalítico no presenta motivos de unión al MeGlcA, de lo que se deduce que el papel de este subdominio en

la unión al sustrato descrito en otras enzimas no es común a todas las enzimas de la familia GH30, particularmente a las modulares.

9. El dominio Xyn30D-CBM35 contiene dos iones calcio, Ca1 y Ca2, con funciones funcional y estructural, respectivamente.
10. Xyn30D-CBM35 presenta dos sitios de unión a GlcA, denominados sitio 1 y sitio 2. Únicamente el sitio 1 tiene relevancia biológica y es capaz de unir tanto GlcA como ácido aldurónico.
11. El residuo Glu31, se reorganiza para formar un enlace de hidrógeno con el O3 de GlcA, por lo tanto cambia su conformación al unir el sustrato.
12. Se ha identificado el residuo Glu129 como clave en la unión presentando un enlace bidentado con el calcio.
13. Se han identificado dos residuos aromáticos, Try34 y Tyr48, como determinantes de unión en la generación de subsitios adicionales.
14. Se ha demostrado el papel esencial del calcio en la unión de GX a Xyn30D-CBM35.
15. El calcio no es el único determinante en la unión de GAX. Xyn30D-CBM35 puede interaccionar de manera inespecífica con otros componentes del xilano, como por ejemplo arabinosa.
16. Xyn30D-CBM35 probablemente dirige la enzima a zonas de la pared celular vegetal que están siendo degradadas, ya que muestra clara especificidad por sustratos provenientes de la degradación de la pectina (GlcA).
17. La enzima presenta un sofisticado mecanismo de degradación de xilano, que se refleja en la diferente especificidad frente a GlcA metilado (más inaccesible) o no metilado (más accesible) de los dominios catalítico y Xyn30D-CBM35 respectivamente.

5. DISCUSIÓN GENERAL

“Me interesa mucho más hablar, o estar en contacto con la gente que piensa lo contrario de lo que yo pienso, que de los que piensan lo mismo que pienso yo”

Salvador Dalí (España 1904-1989)

Pintor, escultor, grabador, escenógrafo y escritor

5. Discusión general

Los polisacáridos son un nutriente importante para numerosos ecosistemas microbianos, por ejemplo para comunidades bacterianas y fúngicas establecidas en el rumen y el intestino grueso de mamíferos, donde juegan un papel importante en la alimentación animal y la salud humana. Asimismo, estos polímeros son de creciente importancia industrial, especialmente en los sectores de bioenergía y bioprocesamiento. Precisamente, la degradación de polisacáridos de pared celular vegetal es un proceso de gran trascendencia y la bioconversión en sus oligosacáridos o monosacáridos constituyentes requiere de la acción de una gran cantidad de enzimas tan diversas como lo son sus sustratos. Un ejemplo de la gran complejidad química y estructural que existe dentro de la pared celular vegetal se refleja en el xilano, componente mayoritario de la hemicelulosa. Su degradación es un proceso muy específico y altamente sincronizado. En nuestro estudio se muestra cómo dos xilanasas de *P. barcinonensis*, Xyn10C y Xyn30D, presentan diferentes especificidades de sustrato en el mecanismo molecular de degradación del xilano. Xyn10C, por su parte, es activa sobre xilanos lineales y ramificados, mientras que Xyn30D únicamente degrada xilanos ramificados y necesariamente sustituidos por metilglucurónico. Esto nos indica que la degradación completa de este polisacárido requiere de la acción coordinada de varias enzimas.

Las enzimas que catalizan la degradación de hidratos de carbono complejos son principalmente GH. Estas enzimas pueden atacar sustratos de difícil acceso y suelen contener CBM. Una actualización reciente de la clasificación de los CBM se ha simplificado en 3 tipos: A, B y C (Apartado 1.4.1) (Gilbert *et al.*, 2013), lo cual ha permitido que se pueda identificar la función de estos módulos según el grupo al cual pertenezcan. Esta clasificación proporciona las bases para la comprensión de la especificidad de estos CBM frente a polímeros tanto aislados como en su entorno natural. Generalmente, la especificidad de unión a carbohidratos de los CBM tiene relación con la actividad y especificidad del módulo catalítico asociado, sin embargo, se reportan cada día más estudios en los que se desvincula esta relación. Estos dominios aumentan y mejoran la actividad de las enzimas de las que forman parte y, aunque el mecanismo es incierto, los CBM probablemente cumplen una función de orientación mediante el aumento de la concentración efectiva de las enzimas en las proximidades del sustrato. Así, en las enzimas del presente estudio, se pueden identificar CBM de las familias CBM9, CBM22 y CBM35. Cada uno de los cuales tiene una especificidad concreta y bien diferenciada del resto. Además, incluso para dominios de la misma familia que se encuentran duplicados dentro de una misma proteína, los determinantes de unión son diferentes.

En el contexto de la gran heterogeneidad de la pared celular vegetal y concretamente en la degradación de xilano, la modularidad es una característica muy común de las enzimas implicadas en dichos procesos. Los CBM presentes en su arquitectura molecular permiten regular su función biológica. Generalmente, se observa que los segmentos conectores entre los distintos

dominios presentan una flexibilidad estructural, lo que puede potenciar o promover la accesibilidad al sustrato. Debido a esta flexibilidad, la cristalización de la proteína completa suele ser un proceso muy complicado y se conocen pocos ejemplos de xilanasas modulares completas caracterizadas estructuralmente mediante cristalografía de rayos X. Xyn30D sería el primer ejemplo de una xilanasas de la familia GH30 bimodular cuya estructura cristalográfica se ha reportado. En el caso de Xyn10C, la gran complejidad de su estructura molecular impidió el estudio cristalográfico de la proteína completa.

Una línea frontera en la investigación de CBM es integrar los conocimientos que se tienen respecto a los dominios aislados en el contexto de la actividad general de las enzimas. Se ha sugerido recientemente la importancia del tamaño y la química de los segmentos conectores que enlazan los CBM con sus módulos catalíticos asociados. Recientemente, gracias a varias técnicas estructurales, ha sido posible avanzar en la caracterización de algunas enzimas modulares. Aún así, el desafío es pensar si es posible predecir el papel exacto de un CBM en el contexto de toda la enzima.

Con relación a Xyn10C, el estudio estructural del dominio N-terminal CBM22-1-CBM22-2, muestra resultados muy interesantes. Por un lado, hay que resaltar la alta flexibilidad interna observada en el tándem, que presenta un amplio rango de conformaciones que adopta CBM22-2 respecto a CBM22-1. Esta característica debe potenciar de forma significativa la orientación de cada CBM con respecto a su sitio de unión. Por el contrario, la naturaleza del segmento conector entre CBM22-2 y el dominio catalítico Xyn10C-CD permite proponer un mecanismo de acción más concertado e integrador. Así, el módulo CBM22-1 anclaría la enzima a un primer sitio de reconocimiento, posibilitando el acceso de CBM22-2 a una amplia zona dentro de la cual se uniría a la región concreta que va a ser degradada por el dominio catalítico Xyn10C-CD. Por lo tanto, es importante tener en cuenta que la naturaleza química del segmento conector entre los distintos módulos es un determinante esencial de especificidad, modulando o limitando la orientación de cada CBM con respecto a sus sustratos específicos (Abbott & van Bueren, 2014).

En lo que respecta a Xyn30D, los resultados estructurales obtenidos nos indican pequeñas diferencias de orientación del módulo catalítico (Xyn30D-CD) respecto del dominio adicional (Xyn30D-CBM35). Esto podría revelar una flexibilidad moderada de la enzima y una actividad concertada entre ambos. Sin embargo, debe tenerse en cuenta que parte del segmento conector está estabilizado mediante enlaces de hidrógeno en el cristal. Por ello, no podemos descartar que en ausencia de restricciones relacionadas con el empaquetamiento cristalino, pudiera existir una mayor flexibilidad conformacional. Además, la diferente especificidad que presentan ambos dominios (Xyn30D-CD requiere de ramificaciones de MeGlcA para la hidrólisis mientras que Xyn30D-CBM35 reconoce sustratos sustituidos con GlcA) da idea de la complejidad los mecanismos responsables del reconocimiento molecular. Es posible que el papel del dominio Xyn30D-CBM35 sea dirigir la enzima a zonas de la pared más

accesibles que se encuentran en degradación. Un apoyo de esta hipótesis es el hecho de que Xyn30D-CBM35 reconoce $\Delta 4,5$ -GalA, compuesto que se produce en la degradación de las pectinas de la matriz que rodea las fibras de celulosa y hemicelulosa en la pared celular vegetal. Una vez unido a estas zonas accesibles, el dominio catalítico entraría en contacto con su sustrato llevando a cabo su función hidrolítica.

Otro aspecto importante para comprender el funcionamiento de un CBM concreto es el estudio de la hidrofobicidad de su superficie. La existencia de plataformas hidrófobas sugieren una interacción estable. En el caso de CBM22-2, incluido en Xyn10C, la existencia de un residuo aromático extra extiende su plataforma hidrofóbica, generando sitios adicionales de unión no descritos hasta el momento para esta familia.

Para el caso de que no se disponga de resultados estructurales, Abbott y colaboradores han propuesto recientemente realizar comparaciones filogenéticas de las secuencias de CBM para identificar patrones que permitan deducir la especificidad de ligando de un CBM. Este enfoque compara directamente la similitud de los residuos funcionales entre secuencias de CBM y resulta ventajoso para agrupar especificidades similares dentro de las familias que unen una diversa variedad de hidratos de carbono. De este modo, para enzimas que contienen múltiples copias de CBM de la misma familia, se han definido patrones homogéneos y heterogéneos. Un patrón homogéneo sugiere mecanismos similares y la repetición implicaría efectos de multivalencia. Un patrón heterogéneo define CBM de la misma familia en diferentes subgrupos lo que indica, probablemente, diferentes mecanismos de unión o distinta afinidad de sustrato. El tándem CBM22-1-CBM22-2 de Xyn10C podría ser un ejemplo del último modelo. Además, el esquema de oligomerización que presenta CBM22-1 en forma de trímero, puede ser un indicativo más de la diferente función que presenta CBM22-1 *vs.* CBM22-2. Aún así, su importancia fisiológica no está clara y se debe investigar más en profundidad este aspecto.

La abundante información de datos bioquímicos y estructurales ha permitido un avance en la comprensión de las bases moleculares de las especificidades mostradas por distintos CBM. Recientemente, se ha propuesto su uso como sondas moleculares, pudiendo utilizar los CBM para determinar la estructura química precisa de las paredes celulares (Gilbert *et al.*, 2013). Nuestra proteína Xyn10C podría ser una buena candidata para ello. Por ejemplo, se podría utilizar la biomolécula completa, que contiene CBM de unión a celulosa (CBM9) y a xilano (CBM22), para identificar zonas de la pared en las cuales celulosa y hemicelulosa estén en estrecha asociación. Del mismo modo, sugerimos que Xyn30D también podría tener una aplicación en este sentido. De este modo, Xyn30D-CBM35, con alta especificidad por GlcA, podría ser utilizado para determinar regiones de la pared que están siendo degradadas, tal como ha sido usado el dominio Pel-CBM35 para reconocer la degradación de la pectina (Figura 5.1) (Montanier *et al.*, 2009).

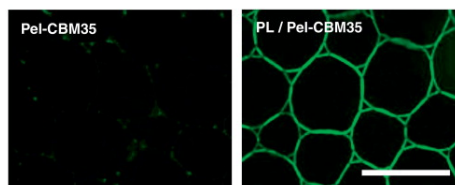


Figura 5.1. La generación de $\Delta 4,5$ -GalA por una pectato liasa (PL) permite la unión de Pel-CBM35 a la pared celular (Montanier *et al.*, 2009). Barra = 10 μ m.

Finalmente hay que resaltar que en base a lo que Hervé y colaboradores observaron (Herve *et al.*, 2010), es posible que cuando existe una arquitectura de dominios compuesta, como es el caso de Xyn10C, la proteína se comporte de forma diferente a si tuviera una composición de dominios simple. La capacidad de las xilanasas de degradar xilano está favorecida tanto por CBM de unión a xilano como a celulosa. Esto es debido a que existen fuertes efectos de proximidad dentro de la pared celular que explican por qué los CBM de unión a celulosa se anexan a muchas hidrolasas de la pared celular no-celulósica.

En conclusión, los análisis estructurales y de especificidad de unión de Xyn10C y Xyn30D presentados en esta tesis proporcionan aspectos elementales y novedosos que ayudan en la comprensión de los mecanismos moleculares de degradación de xilano. La investigación del funcionamiento de este tipo de enzimas modulares es clave para la comprensión de los sofisticados mecanismos que regulan la degradación de la pared celular vegetal. Estas dos enzimas de *Paenibacillus barcinocensis* son reflejo de la complejidad que presenta la degradación del componente mayoritario de la hemicelulosa, el xilano (Figura 5.2). Nuestros resultados han permitido definir determinantes de unión en CBM de las familias 22 y 35 así como establecer sus características funcionales y avanzar en la comprensión de la biología de los mismos. El conocimiento de su bioquímica favorecerá el diseño y producción de nuevos y más eficientes biocatalizadores y permitirá plantear industrias más sostenibles.

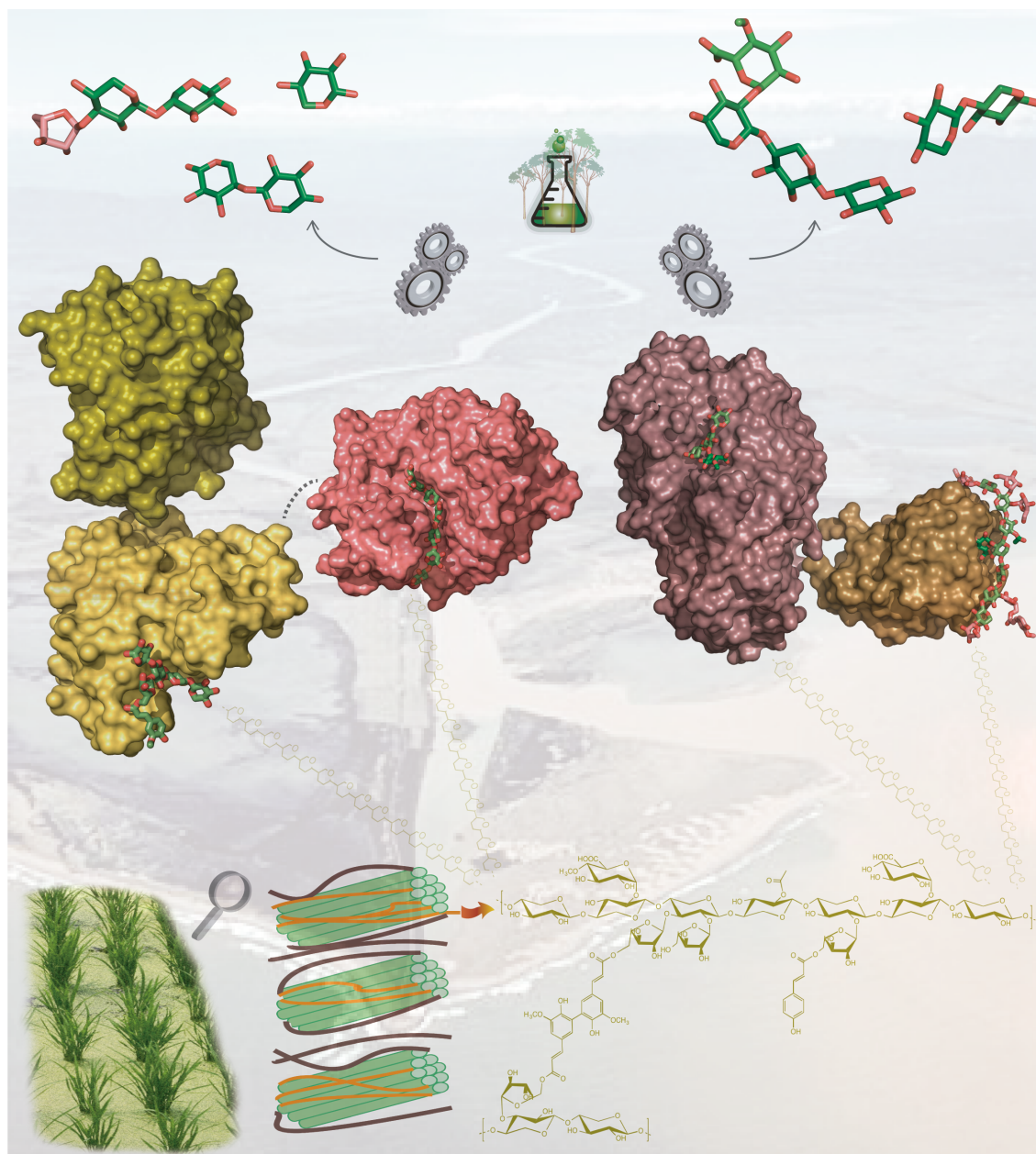


Figura 5.2. Esquema del modo de acción de Xyn10C y Xyn30D en la naturaleza. Xyn10C y Xyn30D han sido aisladas del suelo de un arrozal del delta del Ebro. Son dos xilanasas activas en sustratos que contienen xilano, componente principal de la hemicelulosa presente en la pared celular vegetal. Estas enzimas, asistidas por sus CBM (cada uno de una familia diferente y con distintas especificidades) son capaces de degradar el xilano y generar productos de alto valor añadido. Además, la obtención de estos compuestos por degradación enzimática no genera tantos residuos como la síntesis química tradicional. Por este motivo, tienen una importante aplicación en biotecnología.

6. CONCLUSIONES

*“Un científico debe tomarse la libertad de plantear cualquier cuestión, de dudar
de cualquier afirmación, de corregir errores”*

Julius Robert Oppenheimer (EEUU 1904-1967)

Físico teórico

6. Conclusiones

A continuación se enumeran las principales conclusiones extraídas de esta tesis:

1. Se ha realizado el estudio estructural, mediante cristalografía de rayos X, y funcional de dos xilanasas modulares de *P. barcinonensis*, Xyn10C y Xyn30D.
2. Se ha expresado, purificado y cristalizado el dominio N-terminal de Xyn10C (Xyn10C-XBD), que contiene el tándem CBM22-1-CBM22-2, y se ha resuelto su estructura tridimensional. Este dominio, se pliega en dos módulos independientes, ambos con estructura tipo sándwich β , que están separados mediante un segmento de 10 residuos muy flexible, que confiere una amplia variabilidad conformacional al dominio.
3. La resolución de la estructura del módulo CBM22-1 aislado ha revelado una asociación trimérica similar a la encontrada en Xyn10C-XBD, lo que podría indicar que la oligomerización es una característica de CBM22-1, cuyo papel fisiológico debe ser investigado.
4. Se ha resuelto la estructura de CBM22-2 libre y en complejo con oligosacáridos. El sitio de unión de CBM22-2 es un surco profundo, cuya base está formada por una plataforma hidrofóbica con cierta plasticidad. Destaca el papel del Trp308 como nuevo determinante de unión que define una subfamilia de CBM22 que probablemente une zonas específicas del sustrato.
5. Los estudios de calorimetría y el análisis de geles de afinidad revelan que ambos módulos tienen diferente afinidad frente a oligosacáridos y sustratos, y que la afinidad es mayor en el tándem.
6. La estructura cristalográfica del dominio catalítico Xyn10C-CD revela que es capaz de unir sustratos altamente sustituidos, al igual que se observa en el módulo CBM22-2.
7. Los resultados estructurales y funcionales de Xyn10C han permitido concluir que CBM22-1 podría jugar un primer papel de reconocimiento del sustrato y dirigir la enzima a ciertas zonas del xilano, para permitir que CBM22-2 actúe de forma concertada con el dominio catalítico reconociendo zonas concretas del sustrato. Los dominios de unión a celulosa harían que la enzima pueda estar en proximidad al xilano mediante la unión con las microfibrillas de celulosa.
8. Se ha resuelto la estructura cristalográfica de la proteína Xyn30D completa a 2,4 Å. Se trata de una enzima bimodular que consta de un dominio catalítico con un plegamiento tipo barril unido a una estructura tipo sándwich β de nueve hebras y un dominio CBM35 con un plegamiento tipo lectina.

9. La observación de las ocho moléculas independientes que existen el cristal permite deducir que Xyn30D presenta una flexibilidad moderada en comparación a otras proteínas modulares, aunque no se puede descartar totalmente que exista mayor movilidad en solución.
10. Se ha resuelto la estructura del dominio suplementario Xyn30D-CBM35 y de sus complejos con oligosacáridos que contienen GlcA, lo que ha mostrado que esta sustitución es fundamental para el reconocimiento del sustrato.
11. Se ha observado que el subdominio sándwich β del módulo catalítico de Xyn30D no presenta los sitios de unión a este sustrato. Así, se puede descartar que el papel reportado anteriormente de este subdominio en la unión a sustrato sea una característica común para la familia GH30.
12. Los ensayos de afinidad en geles nativos han mostrado que el calcio es esencial para la unión de Xyn30D-CBM35 a GX. Sin embargo, se ha comprobado que el calcio no es el único determinante en la unión a GAX, ya que Xyn30D-CBM35 es capaz de interaccionar de manera inespecífica con otros componentes del xilano, como podría ser arabinosa.
13. El papel de Xyn30D-CBM35 probablemente es dirigir la enzima a zonas de la pared celular vegetal que están siendo degradadas, ya que muestra clara especificidad por sustratos provenientes de la degradación de la pectina. La diferente especificidad del dominio catalítico por MeGlcA, sustrato metilado y menos accesible, frente a la de Xyn30D-CBM35 por GlcA, sustrato no metilado y más accesible, revela la complejidad del mecanismo molecular de degradación del xilano.
14. Se han descrito nuevos determinantes estructurales responsables de la regulación de la especificidad de Xyn10C y Xyn30D y directamente implicados en la unión del sustrato. Los resultados permiten concluir que para avanzar en la comprensión de las bases moleculares del mecanismo que presentan las enzimas multidominio, es imprescindible combinar el análisis de su dominios aislados con el estudio de la proteína completa.

7. BIBLIOGRAFÍA

“El que lee mucho y anda mucho, ve mucho y sabe mucho”

Miguel de Cervantes Saavedra (España, 1547-1616)

Dramaturgo, poeta y novelista

7. Bibliografía

- Abbott, D. W. & van Bueren, A. L. (2014). *Curr Opin in Struct Biol* **28**, 32-40.
- Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 213-21.
- Alcorn, T. & Juers, D. H. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 366-73.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J Mol Biol* **215**, 403-10.
- Andrews, S. R., Charnock, S. J., Lakey, J. H., Davies, G. J., Claeysens, M., Nerinckx, W., Underwood, M., Sinnott, M. L., Warren, R. A. & Gilbert, H. J. (2000). *J Biol Chem* **275**, 23027-33.
- Araki, R., Ali, M. K., Sakka, M., Kimura, T., Sakka, K. & Ohmiya, K. (2004). *FEBS Lett* **561**, 155-8.
- Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. (2011). *Acta Crystallogr D Biol Crystallogr* **67**, 271-81.
- Bayer, E. A., Shimon, L. J., Shoham, Y. & Lamed, R. (1998). *J Struct Biol* **124**, 221-34.
- Beg, Q. K., Kapoor, M., Mahajan, L. & Hoondal, G. S. (2001). *Appl Microbiol Biotechnol* **56**, 326-38.
- Bellamy, H. D., Snell, E. H., Lovelace, J., Pokross, M. & Borgstahl, G. E. (2000). *Acta Crystallogr D Biol Crystallogr* **56**, 986-95.
- Bergfors, T. (2003). *J Struct Biol* **142**, 66-76.
- Bergfors, T. M. (2009). *Protein crystallization*. La Jolla, Calif.: International University Line.
- Biely, P. (1985). *Trends Biotechnol* **3**, 286-90.
- Biely, P., Vrsanska, M., Tenkanen, M. & Kluepfel, D. (1997). *J Biotechnol* **57**, 151-66.
- Black, G. W., Hazlewood, G. P., Millward-Sadler, S. J., Laurie, J. I. & Gilbert, H. J. (1995). *Biochem J* **307 (Pt 1)**, 191-5.
- Blanco, A., Diaz, P., Zueco, J., Parascandola, P. & Javier Pastor, F. I. (1999). *Microbiology* **145 (Pt 8)**, 2163-70.
- Blanco, A. & Pastor, F. I. J. (1993). *Canadian Journal of Microbiology* **39**, 1162-66.
- Blanco, A., Vidal, T., Colom, J. F. & Pastor, F. I. (1995). *Appl Environ Microbiol* **61**, 4468-70.
- Bolam, D. N., Xie, H., Pell, G., Hogg, D., Galbraith, G., Henrissat, B. & Gilbert, H. J. (2004). *J Biol Chem* **279**, 22953-63.
- Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). *Biochem J* **382**, 769-81.
- Boraston, A. B., Creagh, A. L., Alam, M. M., Kormos, J. M., Tomme, P., Haynes, C. A., Warren, R. A. & Kilburn, D. G. (2001). *Biochemistry* **40**, 6240-7.
- Boraston, A. B., Wang, D. & Burke, R. D. (2006). *J Biol Chem* **281**, 35263-71.
- Brun, E., Moriaud, F., Gans, P., Blackledge, M. J., Barras, F. & Marion, D. (1997). *Biochemistry* **36**, 16074-86.
- Brunger, A. T. (1992). *Nature* **355**, 472-5.
- Burton, R. A. & Fincher, G. B. (2014). *Curr Opin Biotechnol* **26**, 79-84.
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res* **37**, D233-8.
- Carvalho, A. F. A., Neto, P. d. O., da Silva, D. F. & Pastore, G. M. (2013). *Food Res Int* **51**, 75-85.
- Chapman, H. N., Fromme, P., Barty, A., White, T. A., Kirian, R. A., Aquila, A., Hunter, M. S., Schulz, J., DePonte, D. P., Weierstall, U., Doak, R. B., Maia, F. R., Martin, A. V., Schlichting, I., Lomb, L., Coppola, N., Shoeman, R. L., Epp, S. W., Hartmann, R., Rolles, D., et al. (2011). *Nature* **470**, 73-7.
- Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M., Davies, G. J. & Fontes, C. M. (2000). *Biochemistry* **39**, 5013-21.
- Charnock, S. J., Spurway, T. D., Xie, H., Beylot, M. H., Virden, R., Warren, R. A., Hazlewood, G. P. & Gilbert, H. J. (1998). *J Biol Chem* **273**, 32187-99.
- Chávez-Sifontes, M. & Domine, M. E. (2013). *Avances en Ciencias e Ingeniería* **4**, 15-46.
- Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 12-21.
- Clarke, J. H., Davidson, K., Gilbert, H. J., Fontes, C. M. G. A. & Hazlewood, G. P. (1996). *FEMS Microbiol Lett* **139**, 27-35.

- Claus, D. & Berkeley, R. C. W. (1986). In *Bergey's Manual of Systematic Bacteriology* **2**, 1105-39.
- Clemente, J. C., Ursell, L. K., Parfrey, L. W. & Knight, R. (2012). *Cell* **148**, 1258-70.
- Collins, T., Gerday, C. & Feller, G. (2005). *FEMS Microbiol Rev* **29**, 3-23.
- Correia, M. A., Abbott, D. W., Gloster, T. M., Fernandes, V. O., Prates, J. A., Montanier, C., Dumon, C., Williamson, M. P., Tunnicliffe, R. B., Liu, Z., Flint, J. E., Davies, G. J., Henrissat, B., Coutinho, P. M., Fontes, C. M. & Gilbert, H. J. (2010). *Biochemistry* **49**, 6193-205.
- Correia, M. A., Mazumder, K., Bras, J. L., Firbank, S. J., Zhu, Y., Lewis, R. J., York, W. S., Fontes, C. M. & Gilbert, H. J. (2011). *J Biol Chem* **286**, 22510-20.
- Cote, N., Fleury, A., Dumont-Blanchette, E., Fukamizo, T., Mitsutomi, M. & Brzezinski, R. (2006). *Biochem J* **394**, 675-86.
- Coughlan, M. P. & Hazlewood, G. P. (1993). *Biotechnol Appl Biochem* **17 (Pt 3)**, 259-89.
- Cuskin, F., Flint, J. E., Gloster, T. M., Morland, C., Basle, A., Henrissat, B., Coutinho, P. M., Strazzulli, A., Solovyova, A. S., Davies, G. J. & Gilbert, H. J. (2012). *Proc Natl Acad Sci U S A* **109**, 20889-94.
- Dauter, Z. (1999). *Acta Crystallogr D Biol Crystallogr* **55**, 1703-17.
- Davies, G. & Henrissat, B. (1995). *Structure* **3**, 853-9.
- Davies, G. J., Wilson, K. S. & Henrissat, B. (1997). *Biochem J* **321 (Pt 2)**, 557-9.
- de Sanctis, D., Inacio, J. M., Lindley, P. F., de Sa-Nogueira, I. & Bento, I. (2010). *FEBS J* **277**, 4562-74.
- de Vries, R. P. & Visser, J. (2001). *Microbiol Mol Biol Rev* **65**, 497-522, table of contents.
- DeLano, W. L. (2002). *DeLano Scientific.*, San Carlos, CA.
- Delmer, D. P. & Amor, Y. (1995). *Plant Cell* **7**, 987-1000.
- Dhanaraj, D., Byrappa, K., Prasad, V. & Dudley, M. (2010). *Springer*.
- Din, N., Gilkes, N. R., Tekant, B., Miller, R. C., Warren, R. A. J. & Kilburn, D. G. (1991). *Nat Biotech* **9**, 1096-99.
- Ding, S. Y. & Himmel, M. E. (2006). *J Agric Food Chem* **54**, 597-606.
- Ducros, V., Charnock, S. J., Derewenda, U., Derewenda, Z. S., Dauter, Z., Dupont, C., Shareck, F., Morosoli, R., Kluepfel, D. & Davies, G. J. (2000). *J Biol Chem* **275**, 23020-6.
- Emsley, P. & Cowtan, K. (2004). *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32.
- Evans, P. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 72-82.
- Evans, P. R. (2011). *Acta Crystallogr D Biol Crystallogr* **67**, 282-92.
- Festucci-Buselli, R. A., Otoni, W. C. & Joshi, C. P. (2007). *Braz. J. Plant Physiol* **19**, 1-13.
- Ficko-Blean, E., Gregg, K. J., Adams, J. J., Hehemann, J. H., Czjzek, M., Smith, S. P. & Boraston, A. B. (2009). *J Biol Chem* **284**, 9876-84.
- Fontes, C. M., Gilbert, H. J., Hazlewood, G. P., Clarke, J. H., Prates, J. A., McKie, V. A., Nagy, T., Fernandes, T. H. & Ferreira, L. M. (2000). *Microbiology* **146 (Pt 8)**, 1959-67.
- Fry, S. C., York, W. S., Albersheim, P., Darvill, A., Hayashi, T., Joseleau, J. P., Kato, Y., Lorences, E. P., Maclachlan, G. A., Mcneil, M., Mort, A. J., Reid, J. S. G., Seitz, H. U., Selvendran, R. R., Voragen, A. G. J. & White, A. R. (1993). *Physiologia Plantarum* **89**, 1-3.
- Fujimoto, Z., Kaneko, S., Kuno, A., Kobayashi, H., Kusakabe, I. & Mizuno, H. (2004). *J Biol Chem* **279**, 9606-14.
- Gallardo, O., Diaz, P. & Pastor, F. I. (2003). *Appl Microbiol Biotechnol* **61**, 226-33.
- Gallardo, O., Fernandez-Fernandez, M., Valls, C., Valenzuela, S. V., Roncero, M. B., Vidal, T., Diaz, P. & Pastor, F. I. (2010). *Appl Environ Microbiol* **76**, 6290-4.
- Garman, E. (1999). *Acta Crystallogr D Biol Crystallogr* **55**, 1641-53.
- Garman, E. F. & Owen, R. L. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 32-47.
- Georgelis, N., Tabuchi, A., Nikolaidis, N. & Cosgrove, D. J. (2011). *J Biol Chem* **286**, 16814-23.
- Georgelis, N., Yennawar, N. H. & Cosgrove, D. J. (2012). *Proc Natl Acad Sci U S A* **109**, 14830-5.
- Ghosh, A., Luis, A. S., Bras, J. L., Pathaw, N., Chungoo, N. K., Fontes, C. M. & Goyal, A. (2013). *PLoS One* **8**, e80415.
- Gilbert, H. J. (2010). *Plant Physiol* **153**, 444-55.
- Gilbert, H. J. & Hazlewood, G. P. (1993). *J Gen Microbiol* **139**, 187-94.
- Gilbert, H. J., Knox, J. P. & Boraston, A. B. (2013). *Curr Opin Struct Biol* **23**, 669-77.

- Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., Jr. & Warren, R. A. (1991). *Microbiol Rev* **55**, 303-15.
- Gilkes, N. R., Warren, R. A., Miller, R. C., Jr. & Kilburn, D. G. (1988). *J Biol Chem* **263**, 10401-7.
- Gourlay, K., Arantes, V. & Saddler, J. N. (2012). *Biotechnology for Biofuels* **5**.
- Grootaert, C., Van den Abbeele, P., Marzorati, M., Broekaert, W. F., Courtin, C. M., Delcour, J. A., Verstraete, W. & Van de Wiele, T. (2009). *FEMS Microbiol Ecol* **69**, 231-42.
- Hahn, T., Shmueli, U., Wilson, A. J. C. & International Union of Crystallography. (1984). *International tables for crystallography*. Dordrecht, Holland ; Boston, U.S.A.
- Hingham, MA: D. Reidel Pub. Co. ;
- Sold and distributed in the U.S.A. and Canada by Kluwer Academic Publishers Group.
- Hamberg, Y., Ruimy-Israeli, V., Dassa, B., Barak, Y., Lamed, R., Cameron, K., Fontes, C. M., Bayer, E. A. & Fried, D. B. (2014). *PeerJ* **2**, e636.
- Harp, J. M., Timm, D. E. & Bunick, G. J. (1998). *Acta Crystallogr D Biol Crystallogr* **54**, 622-8.
- Hashimoto, H. (2006). *Cell Mol Life Sci* **63**, 2954-67.
- Hassell, A. M., An, G., Bledsoe, R. K., Bynum, J. M., Carter, H. L., 3rd, Deng, S. J., Gampe, R. T., Grisard, T. E., Madauss, K. P., Nolte, R. T., Rocque, W. J., Wang, L., Weaver, K. L., Williams, S. P., Wisely, G. B., Xu, R. & Shewchuk, L. M. (2007). *Acta Crystallogr D Biol Crystallogr* **63**, 72-9.
- Henrissat, B. (1991). *Biochem J* **280 (Pt 2)**, 309-16.
- Heredia, A., Jimenez, A. & Guillen, R. (1995). *Z Lebensm Unters Forsch* **200**, 24-31.
- Hernández-Santoyo, A., Hernández, E. G. & Romero, A. R. (1999). *Journal of the Mexican Chemical Society* **43**, 137-42.
- Herve, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J. & Knox, J. P. (2010). *Proc Natl Acad Sci U S A* **107**, 15293-8.
- Himmel, M. E. & Bayer, E. A. (2009). *Curr Opin Biotechnol* **20**, 316-7.
- Holm, L. & Rosenstrom, P. (2010). *Nucleic Acids Res* **38**, W545-9.
- Holyoak, T., Fenn, T. D., Wilson, M. A., Moulin, A. G., Ringe, D. & Petsko, G. A. (2003). *Acta Crystallogr D Biol Crystallogr* **59**, 2356-8.
- Hooke, R. (1665). *Micrographia*.
- Howlett, G. J., Minton, A. P. & Rivas, G. (2006). *Curr Opin Chem Biol* **10**, 430-6.
- Hurlbert, J. C. & Preston, J. F., 3rd (2001). *J Bacteriol* **183**, 2093-100.
- Jamal-Talabani, S., Boraston, A. B., Turkenburg, J. P., Tarbouriech, N., Ducros, V. M. & Davies, G. J. (2004). *Structure* **12**, 1177-87.
- Jenkins, J., Lo Leggio, L., Harris, G. & Pickersgill, R. (1995). *FEBS Lett* **362**, 281-5.
- Jervis, E. J., Haynes, C. A. & Kilburn, D. G. (1997). *J Biol Chem* **272**, 24016-23.
- Jiang, Z., Dang, W., Yan, Q., Zhai, Q., Li, L. & Kusakabe, I. (2006). *J Biotechnol* **126**, 304-12.
- Kabsch, W. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 125-32.
- Karplus, P. A. & Diederichs, K. (2012). *Science* **336**, 1030-3.
- Kavoosi, M., Meijer, J., Kwan, E., Creagh, A. L., Kilburn, D. G. & Haynes, C. A. (2004). *J Chromatogr B Analyt Technol Biomed Life Sci* **807**, 87-94.
- Kempkes, R., Stofko, E., Lam, K. & Snell, E. H. (2008). *Acta Crystallogr D Biol Crystallogr* **64**, 287-301.
- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958). *Nature* **181**, 662-6.
- Khan, M. I., Sajjad, M., Sadaf, S., Zafar, R., Niazi, U. H. & Akhtar, M. W. (2013). *J Biotechnol* **168**, 403-8.
- Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L. & Woods, R. J. (2008). *J Comput Chem* **29**, 622-55.
- Kolenova, K., Vrsanska, M. & Biely, P. (2006). *J Biotechnol* **121**, 338-45.
- Kraulis, J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J. & Gronenborn, A. M. (1989). *Biochemistry* **28**, 7241-57.
- Krissinel, E. & Henrick, K. (2004). *Acta Crystallogr D Biol Crystallogr* **60**, 2256-68.
- Krissinel, E. & Henrick, K. (2007). *J Mol Biol* **372**, 774-97.
- Kubata, B. K., Takamizawa, K., Kawai, K., Suzuki, T. & Horitsu, H. (1995). *Appl Environ Microbiol* **61**, 1666-8.
- Kulkarni, N., Shendye, A. & Rao, M. (1999). *FEMS Microbiol Rev* **23**, 411-56.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J Appl Crystallogr* **26**, 283-91.

- Laureano-Perez, L., Teymouri, F., Alizadeh, H. & Dale, B. E. (2005). *Appl Biochem Biotechnol* **121-124**, 1081-99.
- Lee, Y. E. & Zeikus, J. G. (1993). *J Gen Microbiol* **139 Pt 6**, 1235-43.
- Linder, M. & Teeri, T. T. (1997). *J Biotechnol* **57**, 15-28.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). *Nucleic Acids Res* **42**, D490-5.
- Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Crystallogr D Biol Crystallogr* **64**, 125-32.
- Luft, J. R. & DeTitta, G. T. (1999). *Acta Crystallogr D Biol Crystallogr* **55**, 988-93.
- Luis, A. S., Venditto, I., Temple, M. J., Rogowski, A., Basle, A., Xue, J., Knox, J. P., Prates, J. A. M., Ferreira, L. M. A., Fontes, C. M. G. A., Najmudin, S. & Gilbert, H. J. (2013). *J Biol Chem* **288**, 4799-809.
- Luo, H., Yang, J., Li, J., Shi, P., Huang, H., Bai, Y., Fan, Y. & Yao, B. (2010). *Appl Microbiol Biotechnol* **86**, 1829-39.
- McCartney, L., Blake, A. W., Flint, J., Bolam, D. N., Boraston, A. B., Gilbert, H. J. & Knox, J. P. (2006). *Proc Natl Acad Sci U S A* **103**, 4765-70.
- McCartney, L., Marcus, S. E. & Knox, J. P. (2005). *J Histochem Cytochem* **53**, 543-6.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J Appl Crystallogr* **40**, 658-74.
- Miyazaki, K., Martin, J. C., Marinsek-Logar, R. & Flint, H. J. (1997). *Anaerobe* **3**, 373-81.
- Montanier, C., van Bueren, A. L., Dumon, C., Flint, J. E., Correia, M. A., Prates, J. A., Firkbank, S. J., Lewis, R. J., Grondin, G. G., Ghinet, M. G., Gloster, T. M., Herve, C., Knox, J. P., Talbot, B. G., Turkenburg, J. P., Kerovuo, J., Brzezinski, R., Fontes, C. M., Davies, G. J., Boraston, A. B., *et al.* (2009). *Proc Natl Acad Sci U S A* **106**, 3065-70.
- Montanier, C. Y., Correia, M. A., Flint, J. E., Zhu, Y., Basle, A., McKee, L. S., Prates, J. A., Polizzi, S. J., Coutinho, P. M., Lewis, R. J., Henrissat, B., Fontes, C. M. & Gilbert, H. J. (2011). *J Biol Chem* **286**, 22499-509.
- Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. (2009). *Acta Crystallogr D Biol Crystallogr* **65**, 1074-80.
- Motta, F. L., Andrade, C. C. P. & Santana, M. H. A. (2013). *A Review of Xylanase Production by the Fermentation of Xylan: Classification, Characterization and Applications*.
- Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Crystallogr D Biol Crystallogr* **67**, 355-67.
- Najmudin, S., Pinheiro, B. A., Prates, J. A., Gilbert, H. J., Romao, M. J. & Fontes, C. M. (2010). *J Struct Biol* **172**, 353-62.
- Nerinckx, W., Broberg, A., Duus, J. O., Ntarima, P., Parolis, L. A., Parolis, H. & Claeysens, M. (2004). *Carbohydr Res* **339**, 1047-60.
- Nishitani, K. & Nevins, D. J. (1991). *J Biol Chem* **266**, 6539-43.
- Notenboom, V., Boraston, A. B., Kilburn, D. G. & Rose, D. R. (2001). *Biochemistry* **40**, 6248-56.
- Notenboom, V., Boraston, A. B., Williams, S. J., Kilburn, D. G. & Rose, D. R. (2002). *Biochemistry* **41**, 4246-54.
- Outtrup, H. & Jørgensen, S. T. (2002). *Edited by R. Berkeley, M. Heyndrickx, N. Logan & P. De Vos. Oxford: Blackwell*, 206-18.
- Pason, P., Kyu, K. L. & Ratanakhanokchai, K. (2006). *Appl Environ Microbiol* **72**, 2483-90.
- Pena, M. J., Zhong, R., Zhou, G. K., Richardson, E. A., O'Neill, M. A., Darvill, A. G., York, W. S. & Ye, Z. H. (2007). *Plant Cell* **19**, 549-63.
- Pereira, N., Couto, M. A. P. G. & Anna, L. M. M. S. (2008). *Brazilian National Library* **2**, 47.
- Pierce, M. M., Raman, C. S. & Nall, B. T. (1999). *Methods* **19**, 213-21.
- Pollet, A., Delcour, J. A. & Courtin, C. M. (2010). *Crit Rev Biotechnol* **30**, 176-91.
- Quiroz-Castañeda, R. E. & Folch-Mallol, J. L. (2011). *Biotechnología Aplicada* **28**, 194-204.
- Raghothama, S., Simpson, P. J., Szabo, L., Nagy, T., Gilbert, H. J. & Williamson, M. P. (2000). *Biochemistry* **39**, 978-84.
- Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J Mol Biol* **7**, 95-99.

- Reyes-Ortiz, V., Heins, R. A., Cheng, G., Kim, E. Y., Vernon, B. C., Elandt, R. B., Adams, P. D., Sale, K. L., Hadi, M. Z., Simmons, B. A., Kent, M. S. & Tullman-Ercek, D. (2013). *Biotechnol Biofuels* **6**, 93.
- Rodriguez, J. M., Murphy, K., Stanton, C., Ross, R. P., Kober, O. I., Juge, N., Avershina, E., Rudi, K., Narbad, A., Jenmalm, M. C., Marchesi, J. R. & Collado, M. C. (2015). *Microb Ecol Health Dis* **26**, 26050.
- Rubinson, K. A., Ladner, J. E., Tordova, M. & Gilliland, G. L. (2000). *Acta Crystallogr D Biol Crystallogr* **56**, 996-1001.
- Sanchez, M. M., Fritze, D., Blanco, A., Sproer, C., Tindall, B. J., Schumann, P., Kroppenstedt, R. M., Diaz, P. & Pastor, F. I. (2005). *Int J Syst Evol Microbiol* **55**, 935-9.
- Saul, F. A., Rovira, P., Boulot, G., Damme, E. J., Peumans, W. J., Truffa-Bachi, P. & Bentley, G. A. (2000). *Structure* **8**, 593-603.
- Scheller, H. V. & Ulvskov, P. (2010). *Annu Rev Plant Biol* **61**, 263-89.
- Schuttelkopf, A. W. & van Aalten, D. M. (2004). *Acta Crystallogr D Biol Crystallogr* **60**, 1355-63.
- Shoseyov, O., Shani, Z. & Levy, I. (2006). *Microbiol Mol Biol Rev* **70**, 283-95.
- Showalter, A. M. (1993). *Plant Cell* **5**, 9-23.
- Simsek, O., Sabanoglu, S., Con, A. H., Karasu, N., Akcelik, M. & Saris, P. E. (2013). *Appl Microbiol Biotechnol* **97**, 4577-87.
- Solomon, V., Teplitsky, A., Shulami, S., Zolotnitsky, G., Shoham, Y. & Shoham, G. (2007). *Acta Crystallogr D Biol Crystallogr* **63**, 845-59.
- St John, F. J., Gonzalez, J. M. & Pozharski, E. (2010). *FEBS Lett* **584**, 4435-41.
- St John, F. J., Hurlbert, J. C., Rice, J. D., Preston, J. F. & Pozharski, E. (2011). *J Mol Biol* **407**, 92-109.
- St John, F. J., Rice, J. D. & Preston, J. F. (2006). *J Bacteriol* **188**, 8617-26.
- Stein, N. (2008). *J Appl Crystallogr* **41**, 641-43.
- Sticklen, M. B. (2008). *Nat Rev Genet*, 433-43.
- Suetake, T., Tsuda, S., Kawabata, S., Miura, K., Iwanaga, S., Hikichi, K., Nitta, K. & Kawano, K. (2000). *J Biol Chem* **275**, 17929-32.
- Sunna, A. (2010). *Appl Microbiol Biotechnol* **86**, 189-200.
- Sunna, A., Gibbs, M. D. & Bergquist, P. L. (2000). *Biochem J* **346 Pt 3**, 583-6.
- Taherzadeh, M. J. & Karimi, K. (2008). *Int J Mol Sci* **9**, 1621-51.
- Taylor, E. J., Goyal, A., Guerreiro, C. I. P. D., Prates, J. A. M., Money, V. A., Ferry, N., Morland, C., Planas, A., Macdonald, J. A., Stick, R. V., Gilbert, H. J., Fontes, C. M. G. A. & Davies, G. J. (2005). *J Biol Chem* **280**, 32761-67.
- Taylor, G. L. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 325-38.
- Timell, T. E. & Syracuse, N. Y. (1967). *Wood Science and Technology* **1**, 45-70.
- Tingaut, P., Eyholzer, C. & Zimmermann (2011). *Functional Polymer Nanocomposite Materials from Microfibrillated Cellulose*.
- Tjalsma, H., Antelmann, H., Jongbloed, J. D., Braun, P. G., Darmon, E., Dorenbos, R., Dubois, J. Y., Westers, H., Zanen, G., Quax, W. J., Kuipers, O. P., Bron, S., Hecker, M. & van Dijk, J. M. (2004). *Microbiol Mol Biol Rev* **68**, 207-33.
- Tomme, P., Boraston, A., McLean, B., Kormos, J., Creagh, A. L., Sturch, K., Gilkes, N. R., Haynes, C. A., Warren, R. A. & Kilburn, D. G. (1998). *J Chromatogr B Biomed Sci Appl* **715**, 283-96.
- Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y. & Steitz, T. A. (1996). *EMBO J* **15**, 5739-51.
- Torronen, A. & Rouvinen, J. (1995). *Biochemistry* **34**, 847-56.
- Urbanikova, L., Vrsanska, M., Morkeberg Krogh, K. B., Hoff, T. & Biely, P. (2011). *FEBS J* **278**, 2105-16.
- Vagin, A. & Teplyakov, A. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 22-5.
- Valenzuela, S. V., Diaz, P. & Javier Pastor, F. I. (2010). *J Agric Food Chem* **58**, 4814-8.
- Valenzuela, S. V., Diaz, P. & Pastor, F. I. (2012). *Appl Environ Microbiol* **78**, 3923-31.
- Van Campenhout, S., Pollet, A., Bourgois, T. M., Rombouts, S., Beaugrand, J., Gebruers, K., De Backer, E., Courtin, C. M., Delcour, J. A. & Volckaert, G. (2007). *Biochem Biophys Res Commun* **356**, 799-804.
- Van Petegem, F., Collins, T., Meuwis, M. A., Gerday, C., Feller, G. & Van Beeumen, J. (2003). *J Biol Chem* **278**, 7531-9.
- Verma, D. & Satyanarayana, T. (2012). *Bioresour Technol* **117**, 360-7.

- Viikari, L., Kantelinen, A., Sundquist, J. & Linko, M. (1994). *Fems Microbiol Rev* **13**, 335-50.
- Vrsanska, M., Kolenova, K., Puchart, V. & Biely, P. (2007). *FEBS J* **274**, 1666-77.
- Wang, J. Y. & Chao, Y. P. (2006). *Appl Environ Microbiol* **72**, 927-31.
- Watson, J. D. & Crick, F. H. (1953). *Nature* **171**, 737-8.
- Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D. & Hochstrasser, D. F. (1999). *Methods Mol Biol* **112**, 531-52.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Crystallogr D Biol Crystallogr* **67**, 235-42.
- Wong, K. K., Tan, L. U. & Saddler, J. N. (1988). *Microbiol Rev* **52**, 305-17.
- Xie, H., Gilbert, H. J., Charnock, S. J., Davies, G. J., Williamson, M. P., Simpson, P. J., Raghothama, S., Fontes, C. M., Dias, F. M., Ferreira, L. M. & Bolam, D. N. (2001). *Biochemistry* **40**, 9167-76.
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. & Zhang, Y. (2015). *Nat Methods* **12**, 7-8.
- Zechel, D. L. & Withers, S. G. (2001). *Curr Opin Chem Biol* **5**, 643-9.
- Zolotnitsky, G., Cogan, U., Adir, N., Solomon, V., Shoham, G. & Shoham, Y. (2004). *Proc Natl Acad Sci U S A* **101**, 11275-80.

8. ANEXOS

*“No tuve educación científica, pero siempre hay un camino para quien quiere
aprender”*

Ada Yonath (Israel, 1939)
Cristalógrafa. Premio Nobel de Química, 2009

| | | | | | | | | | | | | | | | | | | | |
|---|----|----|----|--|--|--|--|--|--|--|--|--|--|--|--|--|--------|----|----|
| 1 | IA | | | | | | | | | | | | | | | | VIII A | | 2 |
| 1 | 1 | 2 | | | | | | | | | | | | | | | | He | |
| | 1 | 2 | | | | | | | | | | | | | | | | Ne | |
| 2 | 3 | 4 | 5 | | | | | | | | | | | | | | | | 10 |
| | 1 | 2 | 3 | | | | | | | | | | | | | | | | Ar |
| 3 | 11 | 12 | 13 | | | | | | | | | | | | | | | | 18 |
| | 1 | 2 | 3 | | | | | | | | | | | | | | | | Kr |
| 4 | 19 | 20 | 21 | | | | | | | | | | | | | | | | 54 |
| | 1 | 2 | 3 | | | | | | | | | | | | | | | | Xe |
| 5 | 37 | 38 | 39 | | | | | | | | | | | | | | | | 86 |
| | 1 | 2 | 3 | | | | | | | | | | | | | | | | Rn |
| 6 | 55 | 56 | 57 | | | | | | | | | | | | | | | | |
| | 1 | 2 | 3 | | | | | | | | | | | | | | | | |
| 7 | 87 | 88 | 89 | | | | | | | | | | | | | | | | |
| | 1 | 2 | 3 | | | | | | | | | | | | | | | | |

Nº ATÓMICO

NOMBRE

6

÷ 4 + 2

VALENCIA

SÍMBOLO

MASA ATÓMICA (uma)

C

CARBONO

12,0107

6

÷ 4 + 2

VALENCIA

SÍMBOLO

MASA ATÓMICA (uma)

C

CARBONO

12,0107

Anexo II – Tabla de aminoácidos

Tabla A.1. Representación de los 20 aminoácidos codificados en el genoma. Se indica su nombre, fórmula y nomenclatura (con los códigos de tres y una letras).

| | | | | |
|--|--|---|--|---|
| $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{CH}_3 \end{array}$ <p>Alanina (Ala / A)</p> | $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{NH} \\ \\ \text{C} = \text{NH} \\ \\ \text{NH}_2 \end{array}$ <p>Arginina (Arg / R)</p> | $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{CH}_2 \\ \\ \text{C} = \text{O} \\ \\ \text{NH}_2 \end{array}$ <p>Asparagina (Asn / N)</p> | $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{CH}_2 \\ \\ \text{C} = \text{O} \\ \\ \text{OH} \end{array}$ <p>Ácido aspártico (Asp / D)</p> | $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{CH}_2 \\ \\ \text{SH} \end{array}$ <p>Cisteína (Cys / C)</p> |
| $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{C} = \text{O} \\ \\ \text{NH}_2 \end{array}$ <p>Glutamina (Gln / Q)</p> | $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{C} = \text{O} \\ \\ \text{OH} \end{array}$ <p>Ácido glutámico (Glu / E)</p> | $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{H} \end{array}$ <p>Glicina (Gly / G)</p> | $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{CH}_2 \\ \\ \text{HN} \quad \text{N} \end{array}$ <p>Histidina (His / H)</p> | $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{HC} - \text{CH}_3 \\ \\ \text{CH}_2 \\ \\ \text{CH}_3 \end{array}$ <p>Isoleucina (Ile / I)</p> |
| $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{HC} - \text{CH}_3 \\ \\ \text{CH}_2 \\ \\ \text{CH}_3 \end{array}$ <p>Leucina (Leu / L)</p> | $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{NH}_2 \end{array}$ <p>Lisina (Lys / K)</p> | $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{S} \\ \\ \text{CH}_3 \end{array}$ <p>Metionina (Met / M)</p> | $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{CH}_2 \\ \\ \text{C}_6\text{H}_5 \end{array}$ <p>Fenilalanina (Phe / F)</p> | $\begin{array}{c} \text{COO}^- \\ \\ \text{C}_5\text{H}_9\text{N}^+ \end{array}$ <p>Prolina (Pro / P)</p> |
| $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{CH}_2 \\ \\ \text{OH} \end{array}$ <p>Serina (Ser / S)</p> | $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{HC} - \text{OH} \\ \\ \text{CH}_3 \end{array}$ <p>Treonina (Thr / T)</p> | $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{CH}_2 \\ \\ \text{C}_8\text{H}_6\text{N}_2 \end{array}$ <p>Triptófano (Trp / W)</p> | $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{CH}_2 \\ \\ \text{C}_6\text{H}_4\text{OH} \end{array}$ <p>Tirosina (Tyr / Y)</p> | $\begin{array}{c} \text{H} \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{COO}^- \\ \\ \text{HC} - \text{CH}_3 \\ \\ \text{CH}_3 \end{array}$ <p>Valina (Val / V)</p> |
| Aminoácidos aromáticos | Aminoácidos polares no cargados | Aminoácidos con carga negativa | Aminoácidos con carga positiva | Aminoácidos no polares alifáticos |

Anexo III – Simetría cristalina

Existen otras operaciones además de las repeticiones por traslación que provocan repeticiones dentro de la celdilla elemental: son los elementos de simetría. Al conjunto de los elementos de simetría de un objeto, los cuales definen la simetría del mismo, se le denomina **grupo puntual de simetría**.

En los cristales, los grupos puntuales deben ser compatibles con la repetitividad por traslación. Teniendo en cuenta esto, no todas las rotaciones son posibles. Así, existen un total de 32 grupos puntuales posibles que se denominan **clases cristalinas**. De estas 32, únicamente 11 contienen el operador centro de simetría y a estas clases cristalinas centrosimétricas se las conoce como **grupos de Laue**. Además, en los cristales, la forma de repetición por traslación necesita ser compatible con la simetría puntal, así que sólo existen 14 tipos de redes de traslación denominadas también **redes de Bravais**.

Al combinar las 32 clases cristalinas con las 14 redes de Bravais, surgen 230 maneras posibles de repetir un “motivo” en el espacio de tres dimensiones: son los **grupos espaciales**. Por último, estas 32 clases, 14 redes y 230 grupos espaciales pueden clasificarse, según la mínima simetría que alojan, en 7 sistemas cristalinos: **triclínico**, **monoclínico**, **ortorrómbico**, **tetragonal**, **cúbico**, **hexagonal** y **trigonal**.

| Sistema cristalino | Ejes | Ángulos |
|--------------------|-------------------|---|
| Triclínico | $a \neq b \neq c$ | $\alpha \neq \beta \neq \gamma$ |
| Monoclínico | $a \neq b \neq c$ | $\alpha = \gamma = 90^\circ$ |
| Ortorrómbico | $a \neq b \neq c$ | $\alpha = \beta = \gamma = 90^\circ$ |
| Tetragonal | $a = b \neq c$ | $\alpha = \beta = \gamma = 90^\circ$ |
| Cúbico | $a = b = c$ | $\alpha = \beta = \gamma = 90^\circ$ |
| Hexagonal | $a = b \neq c$ | $\alpha = \beta = 90^\circ, \gamma = 120^\circ$ |
| Trigonal | $a = b = c$ | $\alpha = \beta = \gamma$ |

“La ciencia y la vida diaria no pueden y no deberían ser separadas”

Rosalind Elsie Franklin (Reino Unido, 1920-1958)

Cristalógrafa y Química. Importantes contribuciones a la comprensión de la estructura del ADN (Watson, Crick y Wilkins compartieron el Premio Nobel de Fisiología y Medicina en 1962. Watson puntualizó que Franklin debió haber sido galardonada también con el Premio Nobel de Química, junto con Wilkins)

“La vida no es fácil para ninguno de nosotros. Pero... ¡qué importa! Hay que perseverar y, sobre todo, tener confianza en uno mismo. Hay que sentirse dotado para realizar alguna cosa y esa cosa hay que alcanzarla, cueste lo que cueste”

Marie Curie (Polonia, 1867-Francia 1934)

Física, matemática y química. Premio Nobel de Física, 1903 y Premio Nobel de Química 1911

